

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 07:37:11 ; Search time 157.11 Seconds

(without alignments)

8332.605 Million cell updates/sec

Title: US-09-604-231-1

Perfect score: 1527

Sequence: 1 ctcatgcatctgcgcgtt.....gttgaaaccttgatgttcg 1527

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_l101.*

```
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	1527	22	AAF31528
2	1315	99.2	34980	22	AAH68533
3	1492	97.7	1983	22	AAH67869
4	1412.6	92.5	5969	22	AAF32543
5	1107.4	72.5	1109	22	AAF31529
6	190.6	12.5	2913	22	AAH54100
7	151.2	9.9	5840	20	AAH12968
8	126.6	8.3	3895	19	AAV52334
9	108	7.1	3615	15	AAQ55752
10	107.8	7.1	9769	19	AAV52163
11	97.8	6.4	465	18	AAAT47502
					C-glutaminc phosph
					C glutamicum codin
					Brevibacterium lac
					C-glutaminc phosph
					S. epidermidis gen
					Enterococcus faeca
					Streptococcus pneu
					Escherichia coli g
					Streptococcus pneu
					Partial P. rubrum

12	86.2	5.6	30246	18	AAV74367	Staphylococcus aur
13	76.4	5.0	465	18	AAV75628	Staphylococcus aur
14	74.2	4.9	474	20	AAH13576	Enterococcus faeca
15	71.6	4.7	357	18	AAAT47505	Partial Enterobact
16	70.4	4.6	9797	20	AAH13487	Enterococcus faeca
17	67	4.4	8494	19	AAV52296	Streptococcus pneu
18	66	4.3	1947	22	AAH81338	Escherichia coli p
19	60.4	4.0	2550	22	AAH54979	S. epidermidis gen
20	60	3.9	6285	20	AAH13352	Enterococcus faeca
21	59.2	3.9	2049	22	AAH68426	C-glutaminc codin
22	59.2	3.9	2172	22	AAH13542	C-glutaminc phosph
23	59.2	3.7	34980	22	AAH68528	C glutamicum codin
24	56.8	3.7	29555	18	AAV74517	Staphylococcus aur
25	54.2	3.5	428	22	AAF31544	C-glutaminc phosph
26	53.2	3.5	2996	22	AAH54445	S. epidermidis gen
27	53.2	3.5	3081	22	AAH54946	S. epidermidis gen
28	53.2	3.5	3932	22	AAH54056	S. epidermidis gen
29	51.8	3.4	7156	20	AAH12966	Enterococcus faeca
30	51.4	3.4	1906	20	AAH13595	Enterococcus faeca
31	51.4	3.4	2882	19	AAV52273	Streptococcus pneu
32	51	3.3	29555	18	AAV74517	Staphylococcus aur
33	49.2	3.2	796	18	AAV74736	Staphylococcus aur
34	45	2.9	3604	20	AAH13417	Enterococcus faeca
35	44.6	2.9	7900	18	AAV74449	Staphylococcus aur
36	44.2	2.9	6092	20	AAH13162	Enterococcus faeca
37	43.2	2.8	760	19	AAZ96397	S. pneumoniae deri
38	42	2.8	400	18	AAV78285	Staphylococcus aur
39	41.6	2.7	341	18	AAV78546	Staphylococcus aur
40	41.4	2.7	10732	21	AAH10594	Gene encoding a su
41	41.2	2.7	6590	20	AAH13048	Enterococcus faeca
42	40.6	2.7	567	21	AAZ29550	HIV codon altered
43	40	2.6	249	20	AAH13801	Enterococcus faeca
44	39.8	2.6	1314	18	AAH84083	DNA encoding a S.
45	39.8	2.6	1314	19	AAV53453	DNA encoding a Pts

ALIGNMENTS

```
RESULT 1
AAF31528
ID AAF31528 standard; DNA; 1527 BP.
XX
AC AAF31528;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutaminc phosphoenolpyruvate DNA #1.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
(BADI ) BASF AG.
XX
Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
WPI; 2001-080989/09.
XX
Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
```

PT for transformation -

PS Claim 3; Page 98-101; 144pp; English.

CC The present invention relates to Corynebacterium glutamicum
 CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
 CC The PTS nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria, the typing or identification of C. glutamicum or
 CC related bacteria, as reference points for mapping C. glutamicum genome,
 CC and as markers for transformation.

XX Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;

Query Match 100.0%; Score 1527; DB 22; Length 1527;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcatggcatctgcgcgttcgcttctccagtggtgtgtttccaccgcaaccaag 60
 DB 1 ctcatggcatctgcgcgttcgcttctccagtggtgtgtttccaccgcaaccaag 60

QY 61 cgtttcggcggaatgagttcctggcgccgctattggtatgcatggtgttcccgag 120
 DB 61 cgtttcggcggaatgagttcctggcgccgctattggtatgcatggtgttcccgag 120

QY 121 cttggtgaacggtacagctggtgcccaccatggtcggcggaatgccaatgtggtc 180
 DB 121 cttggtgaacggtacagctggtgcccaccatggtcggcggaatgccaatgtggtc 180

QY 181 cctgtttgttttagatgttgcccaagccggttacaggccacgtcttctctgtcgtt 240
 DB 181 cctgtttgttttagatgttgcccaagccggttacaggccacgtcttctctgtcgtt 240

QY 241 ggtttcttggaatcttggaacgctgagaaagttcctgcacaagcgactcaagggcactgc 300
 DB 241 ggtttcttggaatcttggaacgctgagaaagttcctgcacaagcgactcaagggcactgc 300

QY 301 agacttccctgatccatccagctgctgacgttgctgctcagcgattcttaccatcctcgc 360
 DB 301 agacttccctgatccatccagctgctgacgttgctgctcagcgattcttaccatcctcgc 360

QY 361 ctttggtcccaagcaatgcgtcgtgggtggcgatgtgctggcacacgctctacaggagcttta 420
 DB 361 ctttggtcccaagcaatgcgtcgtgggtggcgatgtgctggcacacgctctacaggagcttta 420

QY 421 tgattcgtgttccagtcggcggtcgtcctctcgtgctcgttctactcacaatcgtcat 480
 DB 421 tgattcgtgttccagtcggcggtcgtcctctcgtgctcgttctactcacaatcgtcat 480

QY 481 cactggtctgcaccagctcctcccgcaattgagctggagctgtttaaccagggtgagtc 540
 DB 481 cactggtctgcaccagctcctcccgcaattgagctggagctgtttaaccagggtgagtc 540

QY 541 cttcatcttcgcaacggcatctatggtctaataatcgcccagggtgcggcatgtttgscagt 600
 DB 541 cttcatcttcgcaacggcatctatggtctaataatcgcccagggtgcggcatgtttgscagt 600

QY 601 gttcttccctgggaagagtgaaaagctcaaggccttgaggtgcttcagggtgtctccgc 660
 DB 601 gttcttccctgggaagagtgaaaagctcaaggccttgaggtgcttcagggtgtctccgc 660

QY 661 tgttcttggtattacgagagctcgtatcttcggttggaaccttcgctcgtcgtgcgcgtt 720
 DB 661 tgttcttggtattacgagagctcgtatcttcggttggaaccttcgctcgtcgtgcgcgtt 720

QY 721 cttcatcgtatcgttacgcgcagctatcgttggtcgtgtttgattgcaactctttaataatcaa 780
 DB 721 cttcatcgtatcgttacgcgcagctatcgttggtcgtgtttgattgcaactctttaataatcaa 780

QY 781 ggcagttgcgttggcgctgcaggtttcttctggtgttcttattgatgctccagatat 840

DB 781 ggcagttgcgttggcgctgcaggtttcttctggtgttcttattgatgctccagatat 840
 QY 841 ggtcatgtttctggtgtgtgcagttgttaccttcttcatcgcattcggcgagcgattgc 900
 DB 841 ggtcatgtttctggtgtgtgcagttgttaccttcttcatcgcattcggcgagcgattgc 900

QY 901 ttatggcctttacttggttcgcgcgaagcgcagcatgatccagatcaaacgctgtctcc 960
 DB 901 ttatggcctttacttggttcgcgcgaagcgcagcatgatccagatcaaacgctgtctcc 960

QY 961 agtgcctgcaggaaacaccaaagccgaagcagaagcaccgagcagattttcaaacgattc 1020
 DB 961 agtgcctgcaggaaacaccaaagccgaagcagaagcaccgagcagattttcaaacgattc 1020

QY 1021 caccatcatccaggcacctttgaccgtgaagctatgcactgagcagcgtcagcgatgc 1080
 DB 1021 caccatcatccaggcacctttgaccgtgaagctatgcactgagcagcgtcagcgatgc 1080

QY 1081 catgtttccagcggaaagccttgctcggcgttgccatcgtcccaaccaggggcagtt 1140
 DB 1081 catgtttccagcggaaagccttgctcggcgttgccatcgtcccaaccaggggcagtt 1140

QY 1141 agtttctccggtgagtggaagattgtggtggcatttcccatcgtcccatgcttttcgagtt 1200
 DB 1141 agtttctccggtgagtggaagattgtggtggcatttcccatcgtcccatgcttttcgagtt 1200

QY 1201 tcgcacaaagctgaggtatggttccaatgtggatatattgatgcacatttggtttcgcacac 1260
 DB 1201 tcgcacaaagctgaggtatggttccaatgtggatatattgatgcacatttggtttcgcacac 1260

QY 1261 agtaaacctcaacggcagcagcacttaaccgcgtgaaagcagggcgagatgaagccaagc 1320
 DB 1261 agtaaacctcaacggcagcagcacttaaccgcgtgaaagcagggcgagatgaagccaagc 1320

QY 1321 aggggagctgctgtgtaattcgcatttgatgccaattgaagcgtcaggttatgaggttaac 1380
 DB 1321 aggggagctgctgtgtaattcgcatttgatgccaattgaagcgtcaggttatgaggttaac 1380

QY 1381 cagcgcgattgttttcgaattacaagaacacggacccgttaaacacttacggtttggg 1440
 DB 1381 cagcgcgattgttttcgaattacaagaacacggacccgttaaacacttacggtttggg 1440

QY 1441 cgaattgaagcgggagcacaacctgctcaacgtcgcaaaagaaagcgggtgccagcaac 1500
 DB 1441 cgaattgaagcgggagcacaacctgctcaacgtcgcaaaagaaagcgggtgccagcaac 1500

QY 1501 accataagttgaacaccttgagttggttgg 1527
 DB 1501 accataagttgaacaccttgagttggttgg 1527

RESULT 2
 AAH68533/c
 ID AAH68533 standard; DNA; 349980 BP.
 XX
 AC AAH68533;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 7068.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX

Db 781 gtccgaccaaagctgaggtgtccaatgtgatatcttgatcattgatgcacattggttttcgac 840
 QY 1259 acagttaaaactcaacgcgcgaacttaaccgcgtgaagaagcagggcgatgaagtcataa 1318
 Db 841 acagttaaaactcaacgcgcgaacttaaccgcgtgaagaagcagggcgatgaagtcataa 900
 QY 1319 gaaggaggctgctgtgtaattcgatattgatgccattaaagctcaggttatgaggtta 1378
 Db 901 gaaggaggctgctgtgtaattcgatattgatgccattaaagctcaggttatgaggtta 960
 QY 1379 accacgccattgttttcgaattacaagaacacggaccctgtaaacacattacggtttg 1438
 Db 961 accacgccattgttttcgaattacaagaacacggaccctgtaaacacattacggtttg 1020
 QY 1439 ggcgaattgaagcggagccacacctgctcaacgctgcgaagaagaagcggcgccagca 1498
 Db 1021 ggcgaattgaagcggagccacacctgctcaacgctgcgaagaagaagcggcgccagca 1080
 QY 1499 acaccataagttgaaccttgagttgctg 1527
 Db 1081 acaccataagttgaaccttgagttgctg 1109

RESULT 6

AAH54100/C
 ID AAH54100 standard; DNA; 2913 BP.

AC AAH54100;

XX 03-SEP-2001 (first entry)

DT S. epidermidis genomic polynucleotide sequence SEQ ID NO:3464.

DE Staphylococcus epidermidis SRI strain; infection; diagnosis;

KW vaccination; endocarditis; ds.

XX Staphylococcus epidermidis.

OS WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -

PS Claim 8; Page 1030-1031; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to
 CC AAH5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 2913 BP; 1072 A; 480 C; 419 G; 942 T; 0 other;

Query Match 12.5%; Score 190.6; DB 22; Length 2913;
 Best Local Similarity 55.0%; Pred. No. 2.7e-46;
 Matches 468; Conservative 0; Mismatches 364; Indels 19; Gaps 4;

QY 5 tggcatctgcgcgtctctctccagtggttggttttcacgcgaacacacgcgtt 64
 Db 2690 TTGCAATACACCTTTTACATTTATACCAATACTTATTTGTTTGTAGTCAGCTAAGCGAT 2631
 QY 65 tcggcggaatgagttccctggcgccgcgtattggtatgcgatggtgtcccgagcttg 124
 Db 2630 TTGGTGGTAATCCTTTATTTAGGTGCAGC-TCTAGGTATGATACTTGTTCATCCTGGATTG 2572
 QY 125 gtaacggctacgaacgtggcgccaccatggctgcgggcgaa---atgccaatgtggtcc 181
 Db 2571 ATGAGTGCATATGATTTTCCAAAAGCTTTAGAAAGAGGAAAGCTATTTCCACACTGGGAT 2512
 QY 182 ctgtttggttttagatgttgcccaagccggttaccaggccaccgctgtctctgtgctgggtg 241
 Db 2511 GTCCTTTGGGCTACATATTAATGAAGTAGGTATATCAGGGACAAAGTATACCTATGCTCGTA 2452
 QY 242 gttcttgattctggcaacgatcgagaagttcctgcacaagcgcactcaaggcgactgca 301
 Db 2451 GCAACATATATTTTAGCTACGATTGAAAATGTTACGTAAAGTTATTTCCAACTGTGTTA 2392
 QY 302 gacttctgatcaactcagtgctgacgttgctgctcaccgggattctctacattcatcgcc 361
 Db 2391 GATAAATTTATTGACGCCATTTATCAATTTTATTACAGCATTTATAAGCATTTTATTTT 2332
 QY 362 attgcccagcaatcgctgggtggcgatgctggcacacagctctacaggcgacttat 421
 Db 2331 GTAGGCGCTGTCACTCGTCAATTAGGTTATTTGGTTATCTCATGATGATGACTTGGTTATAT 2272
 QY 422 gattcgggtggtccagtcgcggtgctgctctcgtggtgctgctactcaaccaatcgctc 481
 Db 2271 GAATTTGGTGGAGCTATTTGGTGACTTATATTTGGTTTATTTATATATGCGCCAATCGTCATT 2212
 QY 482 actggtctgcaccagctcctcccgccaattgagctg-----gagctgtttaac 529
 Db 2211 ACAGGAATGCACCATAGCTTTATTGCGAGTTGAAACACACATTAATTCGTGATCGACTAAA 2152
 QY 530 cagggtggatcctctcatcttcgaacggcatctatggtaaatatcgccaggggtgcggca 589
 Db 2151 ACAGGTGGTTCATTTATCTTCCCAATCGCAACGATGTCAAAATATTGCACAGGTGGTGCA 2092
 QY 590 tgtttggcagtgcttct---tcctggcgaaagagtgaagaagctcaaggcccttcgagtgct 646
 Db 2091 GCTTTAGCTGCATTCCTTTATCATTAAGCAAAATAAAAAATTAAGAGGTGTGCTTCGCG 2032
 QY 647 tcagggtctcgcgtcttcttggtattacggagcctgcgcatcttcggtgtggaaccttcgc 706
 Db 2031 GCGGGTATTTCAGCTTTACTAGGAATTACAGAACCCAGCAATGTTTGGTGTCAATCTAAA 1972
 QY 707 ctgctgtgcccgttctctcatcggtatcggtacccgcagctatcggtggtgcgtttgatgca 766
 Db 1971 TTGAGATATCCATTTATAGGTGCTGTTCAGGATCAGGTATAGTGTGCGGCTTATATTCA 1912
 QY 767 ctctttaatatcaaggcagttgctggcgctgcaggtttcttgggtgtgttcttatt 826
 Db 1911 TTCTTCAAGTAAAGCGATAGCGCTTGGTACAGCTGGATTACCTTGGATTATATCTATA 1852
 QY 827 gatgctccaga 837
 Db 1851 AATCCTCACACA 1841

[illegible]

Qy	427	cggtggtccagtcggcggtctgctctctcggtctctggtctactcaacaatcgctcatcaactcg 486
Db	2978	AGTTAAATGGCTCTTCGGTGCATAATTTTTTGGCGCCCTCTCAGCTGCCATTTTGTGTCATCAGG 2919
Qy	487	cttgccacagtcctcccgccaattgagctgagctggtt---aaccaaggatggatcctt 543
Db	2918	CTGTCACCATATGACCAATGCCATTGATACAAATTGATTCGGATGCTGTTGGCAGCTGC 2859
Qy	544	catcttcgcaacggcatctatggtctaataatccgccaggtgcggcatgttttgccagtgtt 603
Db	2858	CCATATGGCAATGATTGCTCTTTCTTAATATTGCTCAAGGCTCAGCCGCTGTTTGCTATTA 2799
Qy	604	cttcctggcgaaagagtgaaaagctcaaggccc---ttgcaggtgtctcaggtgtctccgc 660
Db	2798	TTTTCATGCATGCCCATGATGACGTGAGGCTCAGGTTTCACTTCTCTGCAACCAATTTTTCAGC 2739
Qy	661	tgtctctggtattacggagacctgcgatactcggtgtgaaaccttcgcctgcgcgtggccgtt 720
Db	2738	CTATCTCGGTGTTACAGAACCAAGCTCTTTTGGGGTTAAAGTAAATATATTATTCATT 2679
Qy	721	cttcatctggttatcggtaaccgcagctatcggttggcgctttgattgtaacctctttaataca 780
Db	2678	TGTTGTGGGATGACTGTGTTCAAGCCTTCAGGCGATGTTATCCGTTTACTTTTATGTAA 2619
Qy	781	ggcagttgcttggcgctgcaggtttcttgggtgtgttcttcttattcttcagatctccagatat 840
Db	2618	TGGCGGCTCTTATTTGGTATCGGTGGTTGGCAGGTATTCTCTTATTCACCTCAATACAT 2359
Qy	841	ggtcattgtcttggtgtgtgcagttgttacctcttctcattcgcattcggcgagcgattgc 900
Db	2558	GCTGCCATTGTCAGGAACATATGCTAGTTGCGATTGTTGTCATGCTC-----TTGAC 2505
Qy	901	tbtatgaccttacttggttcgcgcgaacggcagcatgatccagatgcgaaccgcctgcctc 960
Db	2504	TTTTCTTTTCCCGCAGGCTGGTCTCTTTACAAAAACAGAGGGCGATACGAACTTGCAAGC 2445
Qy	961	agtgcctcgaggaacgaccaaagccggaagcagaagcaccgcagaaattttcacaacgattc 1020
Db	2444	AGAAATCGTTGCTCAGAGAGAGCAGAGATTGTGAACCATGAACCACTAGACCTTACTTTC 2385
Qy	1021	caccatctcaggcaacctttgaccggtgaagctattgtaactgagcagcgctcagcgatgc 1080
Db	2384	GGTAGAAATATTACGCCCACTAATCTGGCCAAGTAGAAAGAATTAGTCTAAGCGACGGATCC 2325
Qy	1081	catgtttgccagcgaaagcttggctcggcggttgccatcgtcccaaccaaggggcagtt 1140
Db	2324	TATTTTGGCATCAGGTGTCTATGGGCAAGGCTCTAGTCAATTGAACCAAGCCAGGCTGAGTT 2265
Qy	1141	agttctccggtgagtggaaagattgtgtggaattcccatctcgtgcattgctttccgagtt 1200
Db	2264	GACCTCTCCAGTTAATGGGCACAGTACGGTCTCTTTTCCCTACCAAGCATGCCATCGGCAT 2205
Qy	1201	tcgcacaaggctgaggaatggtttccaatgtgatacttctgtagcaacttggttttcgacac 1260
Db	2204	TGTCTCTGACGAGG-----GAGTTGAATTGTGTCATCCACATCCGATATGGATATC 2157
Qy	1261	agtaaacctcaacggcagcgaactttaaccgcgtgaagcagggcgatgaagtcaaacg 1320
Db	2156	AGTAGGTCTTGTATGGCAAGAGGTTTGTAAAGTCTTGTAGTCCAAGGAGATACGTTTACGT 2097
Qy	1321	aggggagctgctgtggaattcgaattgataccatttaagctgcaggttatgaggttaac 1380
Db	2096	TGGTCAGCAACTGATTCGTTTTTGATATGGATGTCTAATGAAGCTGCAGGCTCTGGTGACAGA 2037
Qy	1381	cacgcgagattgttttcgaa 1401
Db	2036	AACTCTGTTTATCATCACCAA 2016

RESULT 9
AAQ55752/c
ID AAQ55752 standard; DNA: 3615 BP.

```

XX AC AAQ55752;
XX DT 25-SEP-1995 (first entry)
XX DE Escherichia coli genomic probe EC-24.
XX KW Probe: S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
XX KW K.pneumoniae; E.cloacae; clinical sample; ds.
XX OS Escherichia coli.
XX PN WO9401583-A.
XX PD 20-JAN-1994.
XX PF 07-JUL-1993; 93WO-JP00936.
XX PR 07-JUL-1992; 92JP-0179719.
XX PA (FUSO ) FUSO PHARM IND LTD.
XX PA (OHNO/) OHNO T.
XX PI Eda S, Matsuhisa A, Ohno T, Uehara H;
XX DR WPI; 1994-035086/04.
XX PT Probe for identifying bacteria causing infectious disease -
XX PT consists of a DNA fragment obtained by HindIII cleavage of the
XX PT pathogenic bacterial genomic DNA
XX PS Claim 7; Page 62-64; 100pp; Japanese.
XX CC The nucleotide sequence of a 3615 bp probe obtained by digestion of
XX CC Escherichia coli genomic DNA with the restriction enzyme HindIII.
XX CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,
XX CC by HindIII digestion of the genomes of Staphylococcus aureus,
XX CC S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
XX CC Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used
XX CC to detect their respective microorganisms in clinical samples.
XX SQ Sequence 3615 BP; 966 A; 962 C; 805 G; 882 T; 0 other;

Query Match 7.1%; Score 108; DB 15; Length 3615;
Best Local Similarity 46.2%; Pred. No. 1.3e-21;
Matches 539; Conservative 0; Mismatches 590; Indels 37; Gaps 4;

QY 312 tcaactcagctgtaagctgtgctgctcaccggtattccattacattcaccgcatcgccattggccag 371
DB 2495 TCACACCATTTGCTATGCTGATGCTATACACCCGCTACCTTTCTGCTGGTGGGCGC 2436

QY 372 caatgcgctgggtggcgatgtctggcacacaggtctacagggactttatgatttcggtg 431
DB 2435 TATCAACCTGGATGAAGGAACTGATTGCGCGCGGTATCTCTGGCTTTATACGCGCGTTC 2376

QY 432 gtccagtccggctgtcgtcttcgtctgtctactcaaccaatcgtcactcactgtctgc 491
DB 2375 CTGCATTTGGCGGCGGTAAATGGCGGCTTCTGGCAATCTTCGTCATGTTGGACTGC 2316

QY 492 accagtcctccgcgaattgagctgagctgtttaccagggtgagcttcattcttcg 551
DB 2315 ACTGGGCGCTGGTGGCGGCTGTATCAATAACTTACCGTGCTGGGTAGCACACCATGA 2256

QY 552 caacggcatctatg---gctaataatgccagggtgcgcatgtttggcagttcttc 608
DB 2255 TCCCGTGTAAATGCCGCAATATATGCGCAGTGGCGGCGCTCGCGCTTCTCTCT 2196

QY 609 tggcgaagatgaaagctcaaggcttgcaggtgtcttcaggtgtctcgcgtgttctt 668
DB 2195 GCGAACGCGATGCGCAGAAAAAGTGGTGGCGGATCAGCGGGCTTCACGAGTCTGTTG 2136

QY 669 gtattacgagcctcgatcttcgtgtgaaccttcgctgcgtgcgtgttccttcacg 728

```

```

Db 2135 GTATCACCGAACCGATATGGCGTCAACCTGCGCGTAAGTACCCCTTTGTTATCG 2076
QY 729 gatacggtaccgagctatcggtgagcgttttgattgacactctttaaataatcaagcagttg 788
Db 2075 CCTGTATCAGTGGGGCTTTGGGGCCACCATATTATGGCTAGCGCAACAAAGTCTACT 2016
QY 789 cgttgccgctgaggtttcttgggtgtgtttctattgatgctccagatattgctcatgt 848
Db 2015 CCTTTGGTTTGGCCAAAGTATTTTACCTTCATGCAAAACCATCCCTGACGGGAATTGATT 1956
QY 849 tcttggtgtgaggtgttaccctttctcattcgcgtcgcgcagcaggttgcattatggcc 908
Db 1955 TCACCGTCTGGCGCAGCGTTATTGGCGGTGTCATTGCCATCGGTTGCGCATTTTTCGGTA 1896
QY 909 ttacttggttcgccgcaacgagcagcattgatccagatgcaacgcgtgctccagtgctg 968
Db 1895 CGGTGATGCTTCAATTTTCATCACCGCTAAAGCTACAGCGGCGAGGTGCCCC----- 1842
QY 969 caggaaacgacaaagccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1028
Db 1843 -----GAAGAGAAAACACCCAGAGGTTATTACACCACTGAGCAGGGCGGTA 1798
QY 1029 tccaggcacctttgaccggtgaagctattgcactgagcagcagcagcagcagcagcagcagc 1088
Db 1797 TCTGTTTACCAGATCAGCGGAGAGATTGTGCTGCTCATTTACGTCGCTGATACCACTGTTG 1738
QY 1089 ccagcgaaagcttgctcggcggttcgcatcgtcccaacaaagggcgagtttctctc 1148
Db 1737 CCAGTGGCCTGTTGGGTAAAGGTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1678
QY 1149 cgtgagtgaaagattggtggtcattcccatctgcctatcgttcgagttcgcacca 1208
Db 1677 CGGTGCGGGTCGAAATGCTTCTGTTTCCGCACATTACA-----CGCCATTTGGCATTTG 1624
QY 1209 aggtgaggtggttccaatgtgatatcttgatgcacattgttgatgcacacagtaaac 1268
Db 1623 AGTCAGATGATGG-----TGTGAGATCTGATTCATGTCGGTATCGACACCGCTAAAC 1570
QY 1269 tcaacggcagcagcactttaaccgctgaagcagcagcagcagcagcagcagcagcagcagc 1328
Db 1569 TGGACGGCAAAATCTTTTCCGCTCACGTCACGTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1510
QY 1329 tgcgtgtgaattcgatattgattgattgattgattgattgattgattgattgattgattgatt 1388
Db 1509 GGCTGATTTCTTTTGTATATCCCTGCTATTTCGCGAGGCGGATTTGATCTGACGACGCGG 1450
QY 1389 ttgttcttcgaattacaagaacacgacctgtataacacttaacgttttgggggaaattg 1448
Db 1449 TATTAATCAGTAATAGCGATGATTTTACGAGCAGTATTACCCACGCGCAGCGCAGATAA 1390
QY 1449 aagcgggagcacaacctgctcaagtc 1474
Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 10
AAV52163/C
ID AAV52163 standard; DNA; 9769 BP.
XX
AC AAV52163;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:30.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.

```



```
XX PN EP751218-A2.
XX PD 02-JAN-1997.
XX PF 27-JUN-1996; 96EP-0110396.
XX PR 28-JUN-1995; 95DE-4023560.
XX PA (SUED-) SUEDEZUCKER AG MANNHEIM/OCHSENFURT.
XX PI Klein K, Mattes R, Stegmaier S;
XX DR WPI; 1997-054676/06.
XX PF
XX PT Cell with reduced sucrose metabolism but with sucrose isomerase
XX PT activity - providing high yields of non-cariogenic sugars, esp.
XX PT trehalulose and palatinose
XX PS Claim 4; Pages 12-13; 26pp; German.
XX CC
XX CC The present sequence comprises a partial DNA sequence from the scr operon
XX CC containing the scrA gene region. The scr operon encodes sucrose
XX CC metabolising enzymes, in particular sucrose isomerase. Cells containing
XX CC at least one DNA sequence encoding a protein with sucrose isomerase
XX CC activity, and having reduced ability to metabolise sucrose, or their
XX CC extracts, optionally in an immobilised form, are used to produce
XX CC non-cariogenic sugars, especially trehalulose and/or palatinose. During
XX CC production of the non-cariogenic sugars, decomposition of sucrose to
XX CC monosaccharides (which are difficult to separate from the product) is
XX CC minimised, so the yield of the disaccharides is improved.
XX SQ Sequence 465 BP; 106 A; 137 C; 149 G; 73 T; 0 other;

Query Match 6.4%; Score 97.8; DB 18; Length 465;
Best Local Similarity 53.5%; Pred. No. 4.8e-19;
Matches 204; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 145 cgcaccatgctgcggcgaaatgcaatgtggtccctgttggtttagattgccc 204
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
441 CGCTGGGCGCTCGGGCGGTTTCCACACCATTGTTCTCGGACTCGATATCGCCAT 382
QY 205 agcgggtaccaggcgacgtgtctctgtgctggtggttcttggattctggcaacgat 264
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
381 GATCGGCTATCAGGTACGGTGTTCGGGTGCTGACGGTGTGTTATGACGTGCT 322
QY 265 cgagaagtctcgcacaagcgactcaaggcgactgcagacttctctgactccactgcagtgc 324
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
321 GGAATAACAGCTCGGAGAGTGTATCCCAACACGCGCTGGACCTGATCTGACGCCATTCT 262
QY 325 gacgttgctgcaccggattctctacattcatcgccattgcccagcaatgcctgggt 384
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
261 GACGGTGGTGATCTCGGCTCTCGCTCAGACGGCTGATGCCACGCGGTGTTGTCGCCGG 202
QY 385 gggcgatgtgctggcacacaggtgtacagggactttatgatttggtggtccagtcggcg 444
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GGGTGACGGCATCTCCTCTGCTGCTCAGACGGCTGATGCCACGCGGTGTTGTCGCCGG 142
QY 445 tctgctcttggtctgtgtactcaaccaatgcatactgctgcacagtccttccc 504
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 ATTCTGTTCGGGGCGCTGTATTCCGCCATCGTTCATCAGCGGATTCACACAGCTTCCA 82
QY 505 gcaattgagctgagctgatt 525
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
81 CGCCGTAGAGGGGGGCTGCT 61

RESULT 12
AAV74367
ID AAV74367 standard; DNA; 30246 BP.
XX
AC AAV74367;
```

```
XX 16-MAR-1999 (first entry)
XX Staphylococcus aureus contig SEQ ID #56.
DE Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX Staphylococcus aureus.
OS
XX Key Location/Qualifiers
FH misc_feature 1741..1800
FT /*tag= a
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3541..3600
FT /*tag= b
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 5341..5400
FT /*tag= c
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 7141..7200
FT /*tag= d
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 8941..9000
FT /*tag= e
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 10741..10800
FT /*tag= f
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 12541..12600
FT /*tag= g
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 14341..14400
FT /*tag= h
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 16141..16200
FT /*tag= i
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 17941..18000
FT /*tag= j
FT /*note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
```


PN EP786519-A2.
 XX 30-JUL-1997.
 XX 07-JAN-1997; 97EP-0100117.
 XX 05-JAN-1996; 96US-0009861.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX WPI; 1997-374922/35.
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 XX
 PS Claim 1; Page 1918; 3271pp; English.
 XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 XX
 SQ Sequence 465 BP; 155 A; 69 C; 88 G; 152 T; 1 other;
 Query Match 5.0%; Score 76.4; DB 18; Length 465;
 Best Local Similarity 54.3%; Pred. No. 1.2e-12;
 Matches 175; Conservative 0; Mismatches 146; Indels 1; Gaps 1;
 QY 581 ggtgcgagcattgttgcagcttcttattattataacaaataagaagtttaaaggtggca 640
 DB 11 ggtgcgagcattgttgcagcttcttattattataacaaataagaagtttaaaggtggca 70
 QY 641 ggtgcgagcattgttgcagcttcttattattataacaaataagaagtttaaaggtggca 700
 DB 71 tctgcgcagcattgttgcagcattcttattattataacaaataagaagtttaaaggtggca 129
 QY 701 ctctgcgcgcgtgcgcgttcttcttcacgttcggtaccgcagcttcggtgcgctttg 760
 DB 130 ttaaaactaagatccattatttgcgcgttcggttcggtaccgcagcttcggtgcgcttat 189
 QY 761 attgcactcttataatcaaggcagttggttgcgcgttcggttcggttcggttcggtt 820
 DB 190 attgctttcttcaggttaagcaatcgcttaggaactgctggttgcgcaggtattatt 249
 QY 821 tctattgagctccacgatattgcttcgttcttcttgcgttcggttcggttcggttccttc 880
 DB 250 tcaatcaatccagcagcaggtggttacactacttcttgcgttcggttcggttcggttccttc 309
 QY 881 gcattgcgcgcagcaggttgcgtt 902
 DB 310 atcattgctataacaggttactt 331

RESULT 14

AAAX13576
 ID AAX13576 standard; DNA; 474 BP.
 XX
 AC AAX13576;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE Enterococcus faecalis genome contig SEQ ID NO:639.
 XX
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO9850555-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08985.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Dillon PJ, Kunsch CA;
 XX
 WPI; 1999-045171/04.
 XX
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 XX
 PS Claim 1; Page 1887; 2084pp; English.
 XX
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 XX
 SQ Sequence 474 BP; 134 A; 85 C; 86 G; 163 T; 6 other;
 Query Match 4.9%; Score 74.2; DB 20; Length 474;
 Best Local Similarity 50.7%; Pred. No. 5.3e-12;
 Matches 178; Conservative 0; Mismatches 173; Indels 0; Gaps 0;
 QY 176 tggctcctgtttgtttagatgttgcacaaagcgggttacaggagccggtcttctctgtg 235
 DB 47 tggcaatattttgtgataccatgttgcaacaaacaaactatgctaccacgaattccggta 106
 QY 236 ctggtgtgttcttggattctggcaacagatcgagaagttctcgcacagcactcaaggcg 295
 DB 107 ttatgtccgctatattcttgcataatttggaataatttttcaataaactctcttca 166
 QY 296 actgcagactctctgatccactcagctgacgttgcgtcgcgcggtactcctacattc 355
 DB 167 tctattgattttacattcacaccattattatccgtaataattactggttcttaacttc 226
 QY 356 atcgccattggcccaagcaatgcgtgggtggcgatgctggtgcacacgctctacaggga 415

Db 227 acagttattggtcccaaatgttattactatctaatgaatcacagatgctattgttgg 286
Qy 416 cttatgatttcggtggtccagtcggtgctgtcttctgttctgtctactcaacaatc 475
Db 287 ttataatgaactgggtctcattaggaatggggatcttgggtgaacatatctactcatt 346
Qy 476 gtcaatcaggtctcaccagctccctccgcgaattgagctgagctgttt 526
Db 347 gttatgactggcttcctcagctcagctatcttcctgcaatcgaaacacagttactt 397

RESULT 15

AAT47505

ID AAT47505 standard; DNA; 357 BP.

XX AC AAT47505;

XX DT 19-FEB-1998 (first entry)

XX DE Partial Enterobacter sp. scrA gene region.

XX KW scr operon; scrA; sucrose metabolising enzyme; sucrose isomerase;
KW reduced metabolism; non-carbogenic sugar; trehalulose; palatinose;
KW decomposition; monosaccharide; improved yield; disaccharide; ds.
XX OS Enterobacter species.

XX PN EP751218-A2.

XX PD 02-JAN-1997.

XX PF 27-JUN-1996; 96EP-0110396.

XX PR 28-JUN-1995; 95DE-4023560.

XX PA (SUBD-) SUEDEZUCKER AG MANNHEIM/OCHSENFURT.

XX PI Klein K, Mattes R, Stegmaier S;

XX DR WPI; 1997-054676/06.

XX Cell with reduced sucrose metabolism but with sucrose isomerase
PT activity - providing high yields of non-carbogenic sugars, esp.
PT trehalulose and palatinose

XX PS Claim 6; Page 14; 26pp; German.

XX CC The present sequence comprises a partial DNA sequence from the scr operon
CC containing the scrA gene region. The scr operon encodes sucrose
CC metabolising enzymes, in particular sucrose isomerase. Cells containing
CC at least one DNA sequence encoding a protein with sucrose isomerase
CC activity, and having reduced ability to metabolise sucrose, or their
CC extracts, optionally in an immobilised form, are used to produce
CC non-carbogenic sugars, especially trehalulose and/or palatinose. During
CC production of the non-carbogenic sugars, decomposition of sucrose to
CC monosaccharides (which are difficult to separate from the product) is
CC minimised, so the yield of the disaccharides is improved.

XX SQ Sequence 357 BP; 46 A; 121 C; 107 G; 81 T; 2 other;

Query Match

Best Local Similarity 4.7%; Score 71.6; DB 18; Length 357;

Matches 152; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 183 tgtttggttttagattgtcccaagcgggttaccaggccacggtcttctgtcgtggtgg 242
Db 67 tcttcggcatcgaagtggcgatgatcggtaccaggccacggtcttcccggtgctgctgg 126
Qy 243 ttcttggatttcggcaacgactcgagaagtctctgcacaagcgactcaagggcactgcag 302
Db 127 cgggtggtttatgacatggtcgaaaaacggtcgccsscggttatccctgacgcgctgg 186

Qy 303 acttctgatactccagtgctgacgttctgtctcaccggtattccttaccattcagcca 362
Db 187 acctgatactccaccogttctgacggtgattatctccggtttatccgctgctgctga 246
Qy 363 ttggcccaagcaatgcgctgggtggcgatgctggcacacacggtctcacagggaactttatg 422
Db 247 tcggcccgccggtcgcgctcgcgacggtatttcgtttatccctcagcacgcttatca 306
Qy 423 atttcggtgtccagtcggcggtctgtctctcttcgttcgttcgttcgttcgttcactc 468
Db 307 gccacgcccggctggctggcggtgctgttcggcggttcgttcgttcgttcactc 352

Search completed: March 22, 2002, 08:53:35
Job time: 4584 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 22, 2002, 06:41:31 ; Search time 2009.21 Seconds
(without alignments)
12337.860 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 ctctgcatctgcgcgtt.....gttgaaaccttgagtggttcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vi: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

FEATURES

Location/Qualifiers

1. .1527

Result No.	Score	Query Match	Length	DB	ID	Description
1	1527	100.0	1527	6	AX069134	AX069134 Sequence
2	1515	99.2	349980	6	AX127152	AX127152 Sequence
3	1492	97.7	1983	6	AX122988	AX122988 Sequence
4	1107.4	72.5	1109	6	AX069136	AX069136 Sequence
5	287.6	18.8	21838	1	PDGRAFOPER	L32093 Pedicoccus
6	287.6	18.8	21839	1	PPSURPOP	L32771 P.pentosace
7	250.8	16.4	5800	1	LL297015	Z97015 Lactococcus
8	231.8	15.2	10264	1	AE004395	AE004395 Vibrio ch
9	230.8	15.1	2508	1	STRSCRA	M2711 Streptococ
10	218.6	14.3	10085	1	AE006222	M76768 Vibrio algi
11	215.2	14.1	2955	1	VIBSCRAX	X69800 S.xylosus s
12	210.6	13.8	2655	1	SXSCRA	Z54445 B.subtilis
13	205.4	13.5	4158	1	BSTREAPR	D83967 Bacillus su
14	201.8	13.2	22197	1	D83967	D86417 Bacillus su
15	201.8	13.2	37900	1	D86417	Z99108 Bacillus su
16	201.8	13.2	208430	1	BSUB0005	AP003365 Staphyloc
17	194.8	12.8	341350	1	AF269422	AF269422 Staphyloc
18	190.6	12.5	2913	1	AF269422	AX144742 Sequence
19	190.6	12.5	2913	6	AX144742	AP001514 Bacillus
20	183	12.0	299850	1	AP001514	AF229829 Pseudomon
21	157.6	10.3	4487	1	AF229829	AP001513 Bacillus
22	146.8	9.6	292550	1	AP001513	AP001508 Bacillus
23	146.4	9.6	296950	1	AP001508	X57401 K.pneumonia
24	143.6	9.4	4621	1	KPSCRYAB	AJ250722 Erwinia a
25	141.8	9.3	6950	1	ERM250722	M81772 Erwinia chr
26	138	9.0	5065	1	ERWBGPA	U34876 Bacillus st
27	136.2	8.9	1398	1	BSU34876	D37921 Alkalophilol
28	135.2	8.9	2819	1	BACISPO	X67750 S.typhimuri
29	134.6	8.8	4322	1	STSCRCOMP	J03006 B.subtilis
30	132.2	8.7	3518	1	BACSACP	X73124 B.subtilis
31	132.2	8.7	97015	1	BSENR	Z99123 Bacillus su
32	132.2	8.7	212150	1	BSUB0020	Y00541 Salmonella
33	130	8.5	1530	1	STSCRA	AL449924 Streptoco
34	126.6	8.3	14515	1	AE007479	AE007479 Streptoco
35	123.4	8.1	151947	2	SPNEU1902	AF206272 Streptoco
36	120.8	7.9	8006	1	AF206272	AF205034 Clostridi
37	117.8	7.7	5600	1	AF205034	AE007557 Clostridi
38	117.8	7.7	10811	1	AE007557	AE007557 Clostridi
39	114.2	7.5	136254	1	ECOUW82	AE007465 Streptoco
40	112.8	7.4	10371	1	AE007465	M15746 E.coli bqlc
41	111.8	7.3	2134	1	ECOBGLC	M16487 E.coli bql
42	111.8	7.3	5270	1	ECOBGLC	AE000449 Escherich
43	111.8	7.3	12989	1	AE000449	AP001509 Bacillus
44	111.2	7.3	300750	1	AP001509	AR012068 Sequence
45	109.6	7.2	3615	6	AR012068	

ALIGNMENTS

RESULT	1
AX069134	
LOCUS	AX069134 1527 bp DNA
DEFINITION	Sequence 1 from Patent WO0102583.
ACCESSION	AX069134
VERSION	AX069134.1 GI:12579016
KEYWORDS	
SOURCE	
ORGANISM	Corynebacterium glutamicum.
	Corynebacterium glutamicum
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
	Actinomycetales; Corynebacterineae; Corynebacteriaceae;
	Corynebacterium.
REFERENCE	1 (bases 1 to 1527)
AUTHORS	Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE	Orynebacterium glutamicum genes encoding phosphoenolpyruvat e:
JOURNAL	sugar phosphotransferase system proteins
	Patent: WO 0102583-A 1 11-JAN-2001;
	BASF AKTIENGESSELLSCHAFT (DE)
FEATURES	Location/Qualifiers
SOURCE	1. .1527

/organism="Corynebacterium glutamicum"

/db_xref="taxon:1718"

101..1507

/note="RXS00315"

/codon_start=1

/transl_table=11

/protein_id="CAC27276.1"

/db_xref="GI:12579017"

/translation="MAWVPSLVNGYDAATMAAGEMPWWSFLGLDVAOAGYGVLP
VLVYSILATIEFLKRLKGTADFLITPVLTLLILGLFLTFIAIGPAMRWGDLVLAHG
LQGLYDFGPGVGLLGLVTSPIVITGLHQSPFPIELFLNQSGSFIATASWANIAG
GAALFAVFLKSEKILKLAGASVAVILGTEPAIFGVNLRWRPFFIGTAAIGG
ALIALFNIRKAVALLAGFLGVSIDAPVMFLVCAVVFIFAFGAIAVGLYLVRRN
GSIDPATAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEALASVSDMAFASKL
GSGVATPTKGOLVSPVSGKIVVAFPSGHAFVTRKAEQDSNVDILMHIGDFVNLNG
THPNLKKQDEKAGELICEFDIDAIRKANGYEVTPIVVSNVKKTKPNTYGLGEIE
AGANLLNVAKEAVPATP"

BASE COUNT 304 a 392 c 430 g 401 t

ORIGIN

Query Match 100.0%; Score 1527; DB 6; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctcatggcatctgcgcgcttgcgcttcttccagatgttggtgtttccaccgcaaccaag 60
DB 1 CTCATGSCATCTGCGCGCTTCGCGTCTTCCAGTGTGTGGTTCCACCGCAACCAAG 60
QY 61 cgtttcggcggaatagtgcttctggcgccgctgattggtgtttccaccgcaaccaag 120
DB 61 CGTTTCGGCGCAATGATGTTCTCGGCGCGCGTATTGGTATGCGGATGTTGCCGAG 120
QY 121 ctgtgtaacgctacacatggtgcgcaccatgctgctgcggcgaaatgccatgtgctc 180
DB 121 CTGGTGAACGGCTACGACGTGGCGCGCCACCATGGCTGCGGCGGAAATGCCAATGTGGTC 180
QY 181 cctgttgggttagatgttcccaagccggttacagggacccgctgttcttctgtcgtggt 240
DB 181 CCTGTTGGTTTAGATCTGCCCAAGCCGTTACCAAGGACCGTGTCTCTGTGCTGGT 240
QY 241 gtttcttgattctgcgaacatcgagaagttctctgcacaagcgactcaaggcgactgc 300
DB 241 GGTTCCTTGGAATTTGCAACGATTCGAGAAGTTCCTGCACAAGCGACTCAAGGGCACTGC 300
QY 301 agacttctactactcagctgctgacgttgcctactcaccgattccttacattcctcgc 360
DB 301 AGACTTCTGATCACTCCAGTGTGACGTTGCTGCTCACCAGGATTCCTTACATTCAATCGC 360
QY 361 cattggcccgcaatgcgctgggtggcgatgtgtgacacaggtctacagggaacttta 420
DB 361 CATTTGGCCAGCAATGCGTGGGTGGCGATGTGCTGGCACACGGTCTACAGGGACTTTA 420
QY 421 tgatttcggttggtccagctgcggcgctgctctcgttgcgttgcgttactcacaacatcgcat 480
DB 421 TGATTTCCGGTGGTCCAGTTCGGCGGCTGCTGCTTCGGGTGCTGTACTACCAATTCGTCAT 480
QY 481 cactggtctcaccagtccttcccgccaattgagctgagctgttttaaccagggtggtatc 540
DB 481 CACTGGTCTGCACCACTCTCCCGCAATGAGCTGGAGCTGTTTAACCAAGGTTGGATC 540
QY 541 ctteattctgcgaacggcatctatggctaataatcgcccgaggtgcgcatgtttggcagt 600
DB 541 CTTCATCTTCGAACGGCATCTATGGCTAATATCGCCCGAGGTCGGCGCATGTTTGGCAGT 600
QY 601 gttcttctggcgaagaagtgaagctcaaggcccttgaggctgcttcacaggtgtctccgc 660
DB 601 GTTCTTCTCGCGAAGAGTGAAGAGCTCAAGGGCTTGCAGGCTGTTCAGGCTGTCTCCGC 660
QY 661 tgttcttggttatcagcagctgcgtatcttgcgtgtgaaaccttcgctgcgtcggcgtt 720
DB 661 TGTCTTGGTATTACGAGGCTGGATCTTCGGTGTGAACCTTCGCGCTGCGCTGGCGCGTT 720

QY 721 ctteatcggtatcggtaccgcagctatcggtggcgctttgattgcactctttaataca 780
DB 721 CTTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACACTTTAATATCAA 780
QY 781 ggcagttgcgttggcgctgcaggtttcttgggtgttgggtttctattgatgtccagatat 840
DB 781 GGCAGTTGCGTGGCGCTGCAGGTTTCTTGGGTGTGTGTTCTATTGATGCTCCAGATAT 840
QY 841 ggtcatcttcttgggtgtgtcagttgttaactcttcttcatcgcattcggcgagcagattgc 900
DB 841 GGTTCATGTTCTTGGTGTGTGAGTTGTACCTTCTTCATCGCATTCGGCGCAGCGATTGC 900
QY 901 ttatggccttacttgggttcgcgcacgagcagcattgatccaatcaaacgctgtctcc 960
DB 901 TTATGGCCTTTACTTGGTTCGCCGCAAGCGAGCATTTGATCCAGATGCAACCGCTGTCTCC 960
QY 961 agtccttcgagggaacgacccaagccgagcagaagcaccgcgcaaatatttcaaacgattc 1020
DB 961 AGTGCCTGTCAGAACGACCAAGCCGAGGAGCAGACCCGCGAGATTTTCAAAACGATTC 1020
QY 1021 caccatcatcagcagcactttgaccggtgaagctattgacactgaagcagcgtccagcagtc 1080
DB 1021 CACCATCATCCAGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCACCGATGC 1080
QY 1081 catgtttgcagcggaagcgttgcctgcggcgttgccatcgctcccaaccgaaggcgagtt 1140
DB 1081 CATGTTTGCAGCGGAAAGCTTGGCTCGGCGTTCGCCATCGTCCCAACCAAGGSGCAGTT 1140
QY 1141 agtttctccggtgagtggaagattgtggtggcaattcccaattccttgccttcttcgagtt 1200
DB 1141 AGTTTCTCCCGTGAAGTGTGGTGGCATTTCCCATCTTGCCATCTTTCGCGAGT 1200
QY 1201 tcgcaccaagcgtgagtgatggttccaatgtggtatcttgcacatgtgtttcgcacac 1260
DB 1201 TCGCACCAGGCTGAGGATGTTCCCAATGTGGATATCTTGTATGCACATTTGGTTCGACAC 1260
QY 1261 agtaacctcaacgagcagcactttaaccgcgtgaaagcagggcgatgaagtcgaagc 1320
DB 1261 AGTAAACCTCAACGGCAGCGCATTAAACCCGCTGAAGAAGCAGGCGCATGAAGTCAAAGC 1320
QY 1321 agggagcgtcgtgtgaattcgatattgattgattgattgattgattgattgattgattgatt 1380
DB 1321 AGGGAGCTCTCTGTGAATTTCGATATTGATGCAATTAAGGCTGCAAGGTTATGAGGTAAC 1380
QY 1381 cagcgcagattgttcttcgaattacaagaaacggacgtgtaaacacttaccggtttggg 1440
DB 1381 CAGCCGATTTGTTTTCGAATTACAAGAAAACCGGACCTGTAAACACTTTACGGTTTGGG 1440
QY 1441 cgaattgaagcgggagcgaacctgtcgaactgcgaagaaagaaagcgggtgccagcaac 1500
DB 1441 CGAAATTTGAAGCGGAGGCCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAAC 1500
QY 1501 accataagttgaacaccttgagtgctcg 1527
DB 1501 ACCATAAGTTGAACCTTGAGTGTTCG 1527

RESULT 2

AX127152/c

LOCUS

AX127152 349980 bp DNA

Sequence 7068 from Patent EP1108790.

AX127152 AX114121

VERSION

AX127152.1 GI:14041140

KEYWORDS

SOURCE

ORGANISM

Corynebacterium glutamicum.

Corynebacterium glutamicum

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Corynebacteriaceae;

Corynebacterium.

1 (bases 1 to 349980)

Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,

Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.

Novel polynucleotides

PAT 11-MAY-2001

JOURNAL Patent: EP 1108790-A 7068 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES Location/Qualifiers
source I. 349980
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
/note="Seq 1 to long (3.309.400) split in 11, seq 7068
2.700.001 3.049.980"
BASE COUNT 81250 a 97718 c 90621 g 80391 t
ORIGIN

Query Match 99.2%; Score 1515; DB 6; Length 349980;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1526; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ctcattgcatctgcgcgttcgctgcttcttccagtgcttgggtttccaccgcaacaaag 60
|||||
Db 111326 CTCATGGCATCTCGCGTTCGGGTTCTTGGCCAGTGTGTTGGTTTACCAGCAACCAAG 111267
|||||

QY 61 cgttcgagcggaatgagttctcttggcgccgcgctattgtatggcgatggtgttcccgag 120
|||||
Db 111266 CGTTTCGGGGCAATGAGTTCCTTGGGCGCG-CTATTGGTATGGCGATGGTGTTCGCCAG 111208
|||||

QY 121 cttggtgaacgctacgactggtgcccaccacatgctgcggcgcaaatcccaatggttc 180
|||||
Db 111207 CTTGGTGAACGGCTACAGCTGCGCCGCCACCATGGCTGCGGGGAAATGCCAATGTGGTC 111148
|||||

QY 181 cctgtttggtttagatgttgccaaagcgggttaccaggccacgctcttccctgtgctggt 240
|||||
Db 111147 CCTGTTGTTAGATGTGTTGCCAAGCCGGTTACAGAGGCACCGTGTCTTCCGTGTGTT 111088
|||||

QY 241 gttttcttgattctggcaacgacatcgagaaagtcttcgcacaagcgactcaaggcgactgc 300
|||||
Db 111087 GPTTCTTGGAATCTGGCAACGATCGAGAAGTTCTCGCAACGCGACTCAAGGGCGACTGC 111028
|||||

QY 301 agacttcttgatcactccagtcgactgctgactgctcaccgattcttaccatctatcgc 360
|||||
Db 111027 AGACTTCTTGATCACCAGTGTGCTGAGTTGCTGCTACCGGATTCCTTACATTCTATCGC 110968
|||||

QY 361 cattggccagcaatgcgctgggtggcgatgtgctggcgacacggtctacagggaacttta 420
|||||
Db 110967 CATTGGCCCAAGCAATCGCTGGTGGCGATGCTGGCACACGGTCTACAGGGGACTTTA 110908
|||||

QY 421 tgattcgtgtgtccagtcggcggtctgctcttgcgtctggtctactaccatctcat 480
|||||
Db 110907 TGATTTCGGTGTGCCAGTCGGCGGTCTGCTCTTCGGTCTGCTACTACCAATCTGTCAT 110848
|||||

QY 481 cactgctgcaccagtccttccgcgaattgagctggaagctgtttaaccagggtgatc 540
|||||
Db 110847 CACTGGTGTGCACAGTCTCTCCGCCCAATTGAGCTGGAGCTGTTTAAACAGGGGTGATC 110788
|||||

QY 541 cttcatcttcgcaacgcatctatgcttaatatcgccagggtgcggcatgtttggcagt 600
|||||
Db 110787 CTTTCATCTTCGCAACGCATCTATGCTAATATCGCCAGGGTGCAGCATGTTGGCAGT 110728
|||||

QY 601 gttcttcctggcgaagagtgaaagctcaaggcccttgcagggtgcttcagggtctccgc 660
|||||
Db 110727 GTTCTTCTCGGGCAAGAGTGAAAGCTCAAGGGCCCTTGCAGGTGCTTCAAGTGTCTCCGC 110668
|||||

QY 661 tgtcttgattaccagaccctcgatctcgtgtgaaaccttcgctcgtcggtgcggt 720
|||||
Db 110667 TGTTCTTGGTATTACCGAGCCTCGCATCTCGGTGTGAACTTCGCGTGCCTGCGTGGCCGTT 110608
|||||

QY 721 cttcatcggtatcggtaccgcagctatcggtggcgcttttgatggcactctttaatcaaa 780
|||||
Db 110607 CTTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATCAA 110548
|||||

QY 781 ggcaggttggtggcgctgcaggtttcttgggtgtgtttcttattgatgtccagatat 840
|||||
Db 110547 GGCAGTGTGCTGGCGCTGCAGGTTCCTTGGGTGTGTTCTTATTGATGATCTCCAGATAT 110488
|||||

QY 841 ggtcatgttcttggtgtgtgcaattgttacccttctcatcgcaattcggcgagcagattgc 900
|||||

Db 110487 GGTCAATGTTCTTGGTGTGTCAGTGTACCTTCTTCATCGCATTCGGCGCAGGATTGC 110428
|||||

QY 901 ttatggcctttacttggttcgccgaacggcagcatttgatccagatgcgaacogctgctcc 960
|||||
Db 110427 TTATGGCCTTTACTTGGTTCGGCGCAACGGCAGCATTTGATCCAGATGCAACCGCTGCTCC 110368
|||||

QY 961 agtgcctgcaggaacgaccaaagccgaagcagaagcaccgcagaaattttcaaacgattc 1020
|||||
Db 110367 AGTGCCTGCAGGAACCAACCAAGCCGAAGCAGAACCCCGCAGAAATTTTCAACCGATTTC 110308
|||||

QY 1021 caccatcatccaggcacctttgaccggtgaagctatttgcaacttcccatctggccatgcttccagtc 1080
|||||
Db 110307 CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGAGCAGCSTCAGCGATGC 110248
|||||

QY 1081 catgtttccagcggaagactggtgctggcgttgcctatgcgtcccaaccaaaggcgagtt 1140
|||||
Db 110247 CATGTTTGGCAGCGGAAGCTTGGCTCGGGCGTGTGCATCTGCCAACCAAGGGGCGAGTT 110188
|||||

QY 1141 agtttctcgggtgagtggaagattgtggtgcaattcccatctggccatgcttccagtc 1200
|||||
Db 110187 AGTTTCTCGGTGAGTGAAGATTCTGGTGGCATTCCTCATCTGGCCATGCTTTTCGAGT 110128
|||||

QY 1201 tcgcaccaagcgtgagatggttcccaatgtgatatcttgatgcacatgtgttccgacac 1260
|||||
Db 110127 TCGCACCAAGCTGAGATGGTTCCTCAATGTGGATATCTTGATGCACATTTGTTTCGACAC 110068
|||||

QY 1261 agtaaacctcaacggcgacgcacttttaacccgctggaagaagcaggcgatgaagtcaaaagc 1320
|||||
Db 110067 AGTAAACCTCAACGGCAGCACACTTTAACCGCTGAAGAAGCAGGGCGATGAAGTCAAAAGC 110008
|||||

QY 1321 aggggagctgctgtggaattcgatattgatgcatttaaggctgcagggttataggttaac 1380
|||||
Db 110007 AGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAAGCTGCAAGTTATGAGTTAAC 109948
|||||

QY 1381 cagcgcgattgtgttccaaattacaagaaacgcgacctgtaaacacttacaggtttggg 1440
|||||
Db 109947 CAGCCCGATTGTTTTCGAATTACAGAAACCCGACCTGTAAACACTTTACGGTTTGGG 109888
|||||

QY 1441 cgaatattgaagcggggcgcaacctgctcaacgctcgcaaaagaaagcgggtgcccagcaac 1500
|||||
Db 109887 CGAAATTGAAGCGGGAGCCAACTGCTCAACGTCGCAAGAAAGAAAGACGGTGCAGCAAC 109828
|||||

QY 1501 accataagttgaaaccttgagttctcg 1527
|||||
Db 109827 ACCATAAGTTGAAACCTTGAGTGTTCG 109801
|||||

RESULT 3
AXI22988 1983 bp DNA PAT 11-MAY-2001
LOCUS Sequence 2904 from Patent EP1108790.
DEFINITION AXI22988
ACCESSION AXI22988
VERSION AXI22988.1 GI:14040476
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 1983)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 2904 20-JUN-2001;
FEATURES KYOWA HAKKO KOGYO CO., LTD. (JP)
source Location/Qualifiers
I. 1983
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 410 a 508 c 558 g 507 t
ORIGIN

Query Match									
Best Local Similarity 97.7%; Score 1492; DB 6; Length 1983;									
Matches 1503; Conservative 0; Mismatches 0; Indels 1; Gaps									
Qy	1	ctcatggcatctgcgcggttcgcggttcttcgcaagtgtgggtttcagcgcaaccaag	60						
Db	481	CTCATGGCACTTCGGCGGTTCCTTGCAGAGTGTGGTGGTTCACGCCAACCAAG	540						
Qy	61	cgtttcggcgcaatgagttcccttggcgcgcggtatttggtatggcgatgggttcccgag	120						
Db	541	CGTTTCGGCGCAATGAGTTCCTGGCGCGG-GTATTGATATGGCGATGGTTCGCCAG	599						
Qy	121	cttggtaacggtacgaacgtggcccccaccatggctggcgcgaaatgccaatgtgct	180						
Db	600	CTTGGTCAACGGCTAGACAGTGGCCGCCACCATGGCTGCGGGCGAAATGCCAATGTGTC	659						
Qy	181	ccgtgttggtttagatgttgcccaagccggtttaccagggaacccgtgcttctctgtgctagt	240						
Db	660	CCGTGTTGGTTTAGATGTTGCCCAAGCCGGTTACCAGGCGACCGTGTCTCTGTGCTCGT	719						
Qy	241	ggttctcttgatctctggcaacgacgcagaaagtctcctgcacaagcgactcaagggcactgc	300						
Db	720	GGTTCTTGGAATCTGSCAACGATCAGAAGTTCCTGCACAAAGCAGCTCAAGGCGACTGC	779						
Qy	301	agacttctgatactccagtcgtagtgcgttgctgtcgtccacggattccttacattcatgc	360						
Db	780	AGACTTCCTGATCAGTCCAGTGTGACGTTGTGCTACCCGATTCCTTACATTCATCCG	839						
Qy	361	caatggcccaagcaatgcgtgggtggcgatgtctggtgcacacggtctacagggaacttta	420						
Db	840	CATTGGCCCCAGCAATGCGTGGGTGGCGGATGTGCTGGCACAGGTCATACAGGGACTTTA	899						
Qy	421	tgatttcgggtggtccagtcggcggtctgctctcgggtctggtctactcaccaatcgctcat	480						
Db	900	TGATTTCGGTGGTCCAGTCGGCGGTCGTGCTCTCGGTCTGCTACTCACCAGTCGTGAT	959						
Qy	481	cactggtctgcacagtccttcccgcgaatcgagctggagctgtttaaccagggtcggtac	540						
Db	960	CACTGTGCTGCACAGTCTTCCCGCAATAGCTGGAGCTGTTTAACAGGGGTGGATC	1019						
Qy	541	cttcaatctcgcaacggcatctatgcttaatatcgcccaaggtgcggcatgtttggcaagt	600						
Db	1020	CTTCATCTTCGAACGGCATCTATGCTTAATATPGCCCCAGGTGCGGCGATGTTGGCAGT	1079						
Qy	601	gttcttctgcgaagagtgaaagctcaaggcccttgcaggtgcttcagggtgtctccgc	660						
Db	1080	GTTCTCTCTGCCAGAGTGAAAGCTCAAGGGCCTTCAGGTGCTTCAGGTGTCCTCGC	1139						
Qy	661	tgttcttggtattcacggagctcgatcttcgggtggaaccttcgctgcgctggcggtt	720						
Db	1140	TGTTCTTGTFATTTACGGAGCCTGCGATCTCGGTGTGAACCTTCGCCCTGCCTGCCGCTT	1199						
Qy	721	cttcaatcggtatcggttaaccgagctatcggtggcgctttgattgcactctttaataca	780						
Db	1200	CTTCATCGGTFATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACCTCTTTAATATCAA	1259						
Qy	781	ggcagttcggttgggcgctgcaggttctctgggtgtgttcttattgaatgctccagatat	840						
Db	1260	GGCAGTTGCGTTGGGCGCTGCAGGTTCTTGGGGTGTTGTTTCTATTGAAGTCTCCAGATAT	1319						
Qy	841	ggtcattgttcttggtgtgtgcagttgttaccttcttcacgtcattcggcgagcgatgtgc	900						
Db	1320	GGTCATGTTCTTGGTGTGTCAGTTGTACCTTCTTCATCGCATTCGGCGCAGCGATTTC	1379						
Qy	901	ttagtgccttacttggttcgcgcgaacggcagcattgatccagatgcgaacgcgtgctcc	960						
Db	1380	TTATGCGCTTTACTTGGTTCGCGCGAACGGCAGCATTCAGATGCAGCCGCTGCTCC	1439						
Qy	961	agtgcctgcaggaacgacaaagccgaagcagaagcaccgcagaaatttcaacagattc	1020						
Db	1440	AGTGCCTGCAGAACGACCAAGCCGAAGCAGAGCAGCCCGCAAAATTTCAACAGATTCT	1499						

QY	1021	caccatcatccaggccacctcttgacccggtgaagctattgcaactgagcagcgtcagcgatgc	1080
Db	1500	CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTGCACTGACGACGGTGCAGCGATGC	1559
QY	1081	catgtttgccacgcgaaagccttgctcgccgctgtccatcgctcccaacaaagggcgactt	1140
Db	1560	CATGTTTGCACGCGGAAAGCTTGGCTCGGGCGGTGCCATCGTCCCAACCAAGGGGCGATT	1619
QY	1141	agttctccggtgagtggaagattggttgccattcccatctggccatgcttttcgcagt	1200
Db	1620	AGTTTCTCCGGTGAGTGGAAGATTGTTGTTGGCATTCCTCATCTGGCCATGCTTTGGCAGT	1679
QY	1201	tcgcaccaagcctgaggtggttccaaatgttgatatcttgatgcacattggtttcgacac	1260
Db	1680	TCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGACACATGTGGTTTGCACAC	1739
QY	1261	agtaaacctcaacgcgcacgacatttaaccctgaagcagcaggcgatgaagtcaaaagc	1320
Db	1740	AGTAACCTCAACGCGACGCACTTTAAACCCGCTGAAGAAGCAGGCGGATGAAGTCAAAAGC	1799
QY	1321	agggagcgtgctgtgtaattcgatattgattgcccattaaaggctgcaggtcttatgcgtaac	1380
Db	1800	AGGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGCAGCTTATGAGGTAAC	1859
QY	1381	cacgcgattgttttgcgaattacaaagaacacgcgacctgtaaacacttcaggtttggg	1440
Db	1860	CACGCGGATTGTGTTTTCGAATTACAAAGAAACCGAGCTGTAAACACTTACGGTTTGGG	1919
QY	1441	cgaattgaacggggagcgaacccctgctcaacgtgcgaagaagacggtgcacgacac	1500
Db	1920	GGAATTGAACGGGGAGGCCACCTGCTCAACGTCGCAAGAAGAAGCGGTGCCAGCAAC	1979
QY	1501	acca 1504	
Db	1980	ACCA 1983	
RESULT	4		
AX069136			
LOCUS	AX069136	1109 bp	DNA
DEFINITION	Sequence 3 from Patent WO0102583.	PAT	25-JAN-2001
ACCESSION	AX069136		
VERSION	AX069136.1	GI:12579018	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Pompejus M., Kroeger, B., Schroeder, H., Zeider, O. and Haberhauer, G.		
TITLE	Corynebacterium glutamicum genes encoding phosphoenolpyruvate e:		
JOURNAL	Patent: WO 0102583-A 3 11-JAN-2001;		
FEATURES			
source	BASF AKTIENGESSELLSCHAFT (DE)		
	Location/Qualifiers		
	1..1109		
	/organism="Corynebacterium glutamicum"		
	/db_xref="taxon:1718"		
CDS	<1..1089		
	/note="FRX00315"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC27277.1"		
	/db_xref="GI:12579018"		
	/translation="YDFGVPVGLLFLGVSPVITGLHQSPFPIELFLNQGGSFIF		
	ATASMANIAGCAACLAFFVLAKESEKLGAGSGAVLGITPAIFGVNLRWRPF		
	IGITGAIIIGALLAENIKAVLGAAGFLGVSDAPDMVFLVCAVPTTFIAFGAAI		
	AVGLYLRNRGSDIDPATAPVPAGTTKAEAPAFESNDSTLIQAPLTGEATLSV		
	SDAPFASGKLGSVAIVPTKQLVSPVSGKIYVAFPSGHFAVTRKADGDSNVDILMH		
	IGMTFVNLNGTHFNPLRKQDQEVKAGELLCFEFDIDAIKAAGYEVTTPIVVSNNKKTGP		
	VNTYGLGETEAGANLNLNKAKEAVPATP"		
BASE COUNT	236 a	278 c	303 g 292 t

ORIGIN

Query Match 72.5%; Score 1107.4; DB 6; Length 1109;
Best Local Similarity 99.9%; Pred. No. 3.7e-289;
Matches 1108; Conservative 0; Mismatches 1; Indels 0;

Qy	419	tatgattcgggtgcacgctgcggcgctctgctcttggcttgcgtctactcaacaaatcgctc	478
Db	1	TATGATTTCGGCGGTTCAGTTCGGCGGTCTGCTCTGGTCTGGTCTACTACCAAAATCGTC	60
Qy	479	atcactggctgcacacgctctctccgcgcaattgagctggagctgtttaaccagagctga	538
Db	61	ATCACTGGTCTGCACCACTCTTCCCGCCAATTTGAGCTGGAGCTGTTTAACCAAGGTGGA	120
Qy	539	tccttcactctgcgcaacgagcatctatggctaaatatacgccagaggtgcgcatgtttggca	598
Db	121	TCCTTTCATCTTCGCAACCGGATCTATGGCTTAATATGCCCAGGGTGGCGATCTTTTGGCA	180
Qy	599	gtgttctctcggcaagagtaaaagctcaaggccttcagagctgttcaggctgtctcc	658
Db	181	GTGTTCTCTTCGGCGAAGAGTGAAGCTCAAGGGCTTCGAGGTGCTTCAGGTGTCCTC	240
Qy	659	gctgtctctggattatacggagcctgcgactctcggtgtgaaaccttcgctgcgctggcg	718
Db	241	GCTGTTCTGGTATTACGGAGCCTCGCATCTCGGTGTGAACCTTCGCTCGCTGGCGG	300
Qy	719	ttcttcacggtatcggtaaccgcaagctatcggctggcgcttgattgacactctttaata	778
Db	301	TTCTTTCATCGGTATCGGTACCGAGCTATCGGTGGCGCTTGGATTGCATCTTTAAATATC	360
Qy	779	aaggcagctgcgttggcgctgcaggtttctctgggtgtgtttctatgatgctccagat	838
Db	361	AAGCGAGTTCGTTGGCGCTGCAGGTTCTTGGTGTGTTTCTATTGATGCTCCAGAT	420
Qy	839	atggtcatgttcttggtgtgtgcagtttgttaccttcttcacttcgcatcggcgagc	898
Db	421	ATGTGTCATGTTCTTGGTGTGTGTCAGTTGTTACCTTCTTCATCGATTCGGCGCAGCGATT	480
Qy	899	gctcatggccttacttggttcgcgcgaacgagcagcattgatccaagtcgaacccgtc	958
Db	481	GCTTATGGCCTTACTTTGGTTCGCCGCAACGGCAGCATTTGATCCAGATGCAACCGTGTCT	540
Qy	959	ccagtgcctgcaggaaacgacaaagccgaagcagaagcaccgcagaaatttcaaacgat	1011
Db	541	CCAGTGGCTGCAGAAAGCAGCAAAAGCCGAAAGCAGAAAGCACCAGCAGAAATTTCAAACGAT	600
Qy	1019	tcaccatcatccaggcacctttgaccggtgaagctattgcactgagcagcgtcaacgat	107
Db	601	TCCACCATCATCCAGGCACCTTTGACCGCGTGAAGCTATTGCACTGAGCAGCGTCACGGAT	650
Qy	1079	gccatgtttgcacgcgaaagccttggctcggcggttgcacatcgctcccaaccaaggcgag	113
Db	661	GCCATGTTTGGCAGCGAAAGCTTGGCTCGGGGTTGCCATCTGCCAACCAAGGGGCGAG	720
Qy	1139	ttaattctccggtgagtgaagattgttggtggcattccactctggccatgctttcgca	119
Db	721	TTAGTTTCTCGGTGAGTGAAGATTGTGGTGGCATTTCCCATCTGGCCATGCTTTTCGCA	780
Qy	1199	gttcgcaccaagcctgaggaatggttccaaatggatatacttgatgcacatgtgtttcgac	125
Db	781	GTTTCGACCAAGGCTGAGGATGGTTCCAAATGTGATATCTTGATGCACATTTGTTTCGAC	840
Qy	1259	acagtaaacctcaacgcgcgacactttaaccgcgtgaagcagcggcgatgaagtcaaa	131
Db	841	ACAGTAAACCTTCAACGCGCAGCAGCTTTTAACCCGCTGAAGAGCAGGCGCATGAAGTCAA	900
Qy	1319	gcaggagcctgctgtgtaattcgatattgatgccattgaagctcgaggttatgaggt	137
Db	901	GCAGGGGAGCTGCTGTGTGAATTCGATTATTGATGCCAATTAAAGGCTGCAGGTTATGAGGTA	960
Qy	1379	accagcgcgattgttgttgcgaattacaagaaacgcgcctgtaaacacttacggtttg	143

Db 961 ACCAGCCGATTTGTTCGAATTACAGAATAACCGGACCTGTAAACACTTACGGTTTG 1020

Qy 1439 ggcgaattgaagcggagccaaactgctcaacgtcgcaagaagaagcggtagcacga 1498
|||||
Db 1021 GCGAATTTGAGCGGGAGCCCAACTGCTCACGTCGCNAAGAAAGACGGTGCCACGA 1080

Qy 1499 acaccataagttaaacccttgagtgttcg 1527
|||||
Db 1081 ACACCATAAGTTGAACCTTGAGTGTTCG 1109

RESULT 5

PDCRAFOPER/c

LOCUS PDCRAFOPER 21838 bp DNA BCT 23-MAY-1994

DEFINITION Pediococcus pentosaceus raffinose operon genes.

ACCESSION L32093

VERSION L32093.1 GI:493181

KEYWORDS agar gene; agsS gene; agl gene; alpha-galactosidase;
alpha-glucosidase; fructokinase; insertion element; permease; rafP
gene; rafa gene; raffinose operon; regulatory protein; scrA gene;
scrB gene; scrK gene; scrR gene; sucrose-6-phosphate;
transport protein.

SOURCE
Pediococcus pentosaceus (strain PPel.0) DNA; Insertion sequence
IS30 homolog (transposable element insertion sequence IS30 homolog,
kingdom Prokaryotae) DNA; Pediococcus pentosaceus (strain PPel.0)
DNA; Insertion sequence IS3 homolog (transposable element insertion
sequence IS3 homolog, kingdom Prokaryotae) DNA; and Pediococcus
pentosaceus (strain PPel.0) DNA.

ORGANISM
Pediococcus pentosaceus
Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
Pediococcus.

REFERENCE
1 (bases 1 to 21838)

AUTHORS Leenhouts, K.J., Bolhuis, A.A., Kok, J.J. and Venema, G.G.

TITLE The sucrose and raffinose operons of *Pediococcus pentosaceus* PPel.0

JOURNAL Unpublished

COMMENT On May 25, 1994 this sequence version replaced gi:475106.
FEATURES
source
1..21838
Location/Qualifiers
/organism="Pediococcus pentosaceus"
/strain="PPel.0"
/db_xref="taxon:1255"
487..510
/note="DR2; putative"
/rpt_type=direct
488..561
/note="IRL; putative"
/rpt_type=inverted
513..561
/note="DRI; putative"
/rpt_type=direct
2724..2760
/standard_name="terminator 4"
/note="putative"
complement(2929)..3864)
/gene="rafr"
complement(2929)..3762)
/gene="rafs"
/function="regulation raffinose-operon"
/note="putative"
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="AAA25562.1"
/db_xref="GI:475107"

/translation="MNGEYKTVLANKSFESNVLFFGQBACLPNYYTKGNVRDSVIHYH
IQEGKGTFAAAHNPATVILKKAGDIFLPKGTFCFQADNDQPMKFVGFGSAGIRIAM
LSGSLLAQCKLVROGVNIYADLSELYKVLIHLINDVLLGSLIVRLFYDLIARTSP
PADATNTIKVKSTEOFNLAVSYLQENYSTGTCTMDLCHVNLRSRYLTFLFKTHANTSP
QKLITKRLLEDAKQRLETSNNNSVQSIAIMWGYKDSFTFSKAFFKYSASPSYRKSGIG
1"

-10_signal complement(3837)..3842)
/gene="rafr"

-35_signal complement(3859)..3864)
/gene="rafr"

gene	/gene="rafr" 3909..5918 /gene="rafp" 3909..3914 /gene="rafp" /note="putative" 3932..3937 /gene="rafp" 3993..5918 /gene="rafp" /function="raffinose transport protein" /note="putative" /codon_start=1 /transl_table=11 /product="permease" /protein_id="AAA25563.1" /db_xref="GI:475108" /translation="MQEENHKWGRGIYFGAKGNDAFYSLSYLIFITSHLFD TGKALNDRMVSIVTLIIIMVLRIVLELIDPFIIGNAIDRTKNSGHPFPMVVGTVSS IILLFTNLGGIYAKNMIYLVFAILYITMDIFYSKDVGFWMLPSLITDSRERE KTAFARLSTIGGLVGLVMPAVIFFSAKATSTGDNRGWTFALIICLIALISAWG VGLTREVSDIRKQDVTGVMEIEKALAKNDOLLAAALYLFYGVGNILGSLVY YFYIMGKPKSILSIINIEFLGIATSLFVLVLSKFSKRGVFAAGLVFMGLAIFT IAGSNLWLLAATMGFFGQQMVLVLMVITDSVEYQGLKIGHRDESIALSVRPLID KFGGALSNGVVGQIAIISGMTGTGATSSITAAQLHFKLTMFAFPALLLIIAGIESK QIFLTEKNGAIEVAELERTWRTKFDNTIDQAEKVVTSILDATPIAGQVPIAQVNDP TFAAGTLGDGFALKPSDGRILAPEDATVQVETTRHAGVLGVDGIVLIIHIGLGVK LRTGQFISYVEEGHVQOGDELLEFWDPTIKOAGLDDTVIMTVNSTETETMDWLKVP QAVKATDNIILQHTRA" 5976..8177 /gene="agaR" 5976..8177 /gene="agaR" /codon_start=1 /transl_table=11 /evidence="experimental" /product="alpha-galactosidase" /protein_id="AAA25564.1" /db_xref="GI:475109" /translation="MSLIIVDQANRVFHLHNQTLSTVIFAVEQGGTSLHLYFGHVDHY HGELRPVDRGSGNLPSTDRFSRDLPREYSTAGENDYHLPAAIVRHDTGANA YLVYQGYRIAGKPLGSLPAAPVEDETEAETLIVLVDQSOVEFDLQYTIYDRPV VTRSVQVQCHQHAVNLKERVASMQIDFTDQFETITLPGAHANERHPERGSINIGIQT FSLRGTSSHQNNPFLALVDHTTTFESGDAYGNLVYSGNHAPELEKDDOLDHLAVG INSYENWOLKACATQETPEVLMVYTNKGLNAMOAYHHLIRERVVRSEKNOERPIV VNNWEATEFDENEAKLPIVDEAKOLGIEFVLDLDCWFGCHRDDNSSLGSDWQVDRHKP PQGLNHFVYIEHOGKFGIWLPEMISDSKLYIQHPDILMGVGRGSPSPRQYIIL DLGRQVRNIEFDQLDLKSKOIDIYIKWDMNRHLSDIYSVALPPEQGEVYHRVYLG LYELLERLTATYPHILFGCGSGGGRFDAGMAYMPQIWAISDNTDAVARLTIQYGTSL AVPISLATAHVSPNQOQRETSMSTRSAVAASGVLYGELDLTQLSSADKOIVQKQV VOYKOLRPLIQGEERYELKSPITSNOAAMVSPQODEAIVMVENLTSVAQPSLTKTK LVGLNPKLNQNIATKAIFFGDELMLQGFDPVYQDYTKYVHFKAIVEN" 8265..8289 /standard_name="terminator 3" /note="putative" 9403..9426 /note="DR2; putative" /rpt_type=direct 10396..10442 /note="DR1; putative" /rpt_type=direct 10397..10443 /note="IR2; putative" /rpt_type=inverted 10681..10716 /standard_name="terminator 2" /note="putative" complement(10727..11593) /gene="scrK" CDS complement(10727..11593) /gene="scrK" /function="fructose 6-phosphotransferase"	gene	/note="putative" /codon_start=1 /transl_table=11 /product="fructokinase" /protein_id="AA225565.1" /db_xref="GI:475110" /translation="MLLGAIEAGGTFKVCATGAENGQVSDRISPIPTTPTVETWAVDD YFTTHPDAIGISGFGPIGVNPHDPKRYGITTTPKPGWGDGDFLGHKLSQFNIPLYWT TDVNEAAYGESPMIGIAKDPVNSIYMTIGTGVAGVTSQNHIFNGRHTHELGHMLNRL PGDFFKSNCPYHDICLEGAAAGPAVKRTGKAGKDI PVDDPVWPIITDYTAQACVNL VAFAPDKIILNGSVNMQRQLFPNIREKFAAYLNGYEVPPLDDYIYPAGLGNNSGIAG GLLTAQAALKNA" complement(11673..13832) /gene="agaS" CDS complement(11673..13832) /gene="agaS" /note="putative" /codon_start=1 /transl_table=11 /product="alpha-galactosidase" /protein_id="AAA25566.1" /db_xref="GI:475111" /translation="MPVEYDPKTLNLHNDQISYVIOILLAHXPVHRVGRGYFSKOP YFEPMSGHAFANDPTEPPYVTSLSPEYSTIGSDYRQPAIYKIDANNQLLPIL YTGFSVNDQPIINSRQLPPTVSKHPTVTILVHLTDAVTKLQMDLNTYIFENQQLPIL TLRHHGTNLQVLTALSSAQLDLPTDQYATLTLSTGTHAEANFSNRLHPGLQTVRS RTRSGQHOPFMAALAPNTLAGTIGCALAWSNFDSTVEVDQYQHSRLTIGLEPD TEWOLKPNSSFOTPEAVLTWTNGFNGSOVPHDFSYOLMPSQTNIPSVLNTWETLT FAVESKVLHLEHAHQLOMLVLDGFWNVRNGENGQLGDWFDVPIKFPNGLNPLA QAAHHHRMFGLWVEPEMITNSLYQQHPVWLYQVYDPTITARHOLVLDLSQAAR DHLITTLNLQNNDQYLDKWMNRHLTVQGSTHLPAAQGGELYHRYVCGLYDILTRL KRACPKLIIENG-SAGGRDFGMLNTQNTWISDLTPDVRATIEGFSYLPFPPIFS NHTASPNAGNRIITPETFRLQALCIGQLGLEUNLPOLAPSEQQLLRGALIKYQOLKS TRKAHFYRLPTRHVAVMLVTADKKQAICCVLNGLSNRVKTQHPLPLHYLDAELAY SDSSGNRYTGHQNLTMGIPKLPNADFTSQLIYLCN" complement(13852..15989) /gene="scrA" CDS complement(13852..15807) /gene="scrA" /function="enzyme II of the PTS system, sucrose specific" /note="enzyme IIabc" /transl_table=11 /evidence="experimental" /protein_id="AAA25567.1" /db_xref="GI:475112" /translation="MNHQEVADRVLAIGKNNITQAAAHCAATRLRLVVKDESKIDQAL DDADVKGFETNGQYQIIIGPGVDKVDYDALIVTGLKEVTPDDIKAVAAAGQKNP LMDFLKVLSDIPIYPAIVAGLLMALNNVLTAEHLFMAKSVVEYVPGLKGAEMIN AMASAPFTLPIILGFSATKRFNGPYLGATGMIMVLPVLNGVSVYVATTMAAGRMVY WNVFGLHVAQAGVQGOVLPVLGVATLATEKFFHKIKGAFDFTPMFAIVITGFL TTIVGVLRTVSDALTNGLVLYNSTGWTGICGLLYSALVITGLHOTFPALITQEP LANVATGGSFIPFVNASMANIGGAATLAIFFATKSKOKALITSSAGCVSALLGITEPA IFGNLKMRFPEFVAASIASIAFLGFLHVSVMGPASVIGFITSKASIPAFMLS AVISFVAFIPTFIYAKRTLGDQDQKSPATSVINYNDEIISAPVTSAGESLQKV NDQVFAEIMGKAAIIPSSDQVAPADGVITVYDSHAYGIKTTAGAEILHLGLD TVNLNGEHFTTNVOKGDTVHOGDLTGTDIAALKAAANYDPTVNLIVTNTANVANVERL KVTNQAGEQLVATAPAASSVAATTV" complement(15872..15877) -10_signal /gene="scrA" -35_signal complement(15894..15989) /gene="scrA" -35_signal 15993..15998 /gene="scrB" 15993..17568 /gene="scrB" -10_signal 16017..16022	terminator	Query Match 18.8%; Score 287.6; DB 1; Length 21838; Best Local Similarity 58.8%; Pred. No. 8e-67; Matches 539; Conservative 0; Mismatches 364; Indels 13; Gaps 2;
------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------


```

/transl_table=11
/product="permease"
/protein_id="CAA83664.1"
/db_xref="GI:475964"
/db_xref="SWISS-PROT:P43466"
/translation="MOEHNYKWWGRLIYGFAGKGNDAFYSLSGYLIFITSHLFD
TGKALIDNRMSLVTLIMVLRIVELFIDPFIGNAIDRTKSPGHPRPVWVGTVSS
IILLFLTNGLYAKNAMLYLVFAILYITMDIFYSKDVGFWMLPSLTJDSRERE
KTATFARLSTIGGLVGLVMPAVIFFESAKATSTGDNRCWFELIICLIALISAWG
VGLTREVSDIRKKNQDQIVGYMEIFKALAKNDQLLWAAALVLYFGVGNILGSLVY
YFTYIMGPKPSILSINIIFLGIATSLFVLSKRSKRGVAGCLVFMJGGIAIFT
IAGSNLWLLAATMFGFPQWVFLVLMVITDSVEYQGLKGRHDSLAVRPLID
KFGAISNGVVGQIAIISGTTGATASSITAAQQLHFKLTMAFPALMLLIAIGIFSK
QIFTEKHAEIIVAELEWRTKFDNTTQVAKKVTSLDLATPIAGQVLIPLAQNVP
TPAAGTLGDGAFTKPSDGRILAPEDATVQVETTRHAGVLGVDNGIYLHILGLCTVK
LKGTFISYVEGQHVQGDDELLEFWDPTIKQGLDDTIVIMTVNSTEFTMDMLVKP
GQAVKATDNILQLHTKA"
gene
5976..8177
/genes="agar"
CDS
5976..8177
/genes="agar"
/standard_name="alpha-galactosidase"
/citation=[1]
/codon_start=1
/transl_table=11
/evidence=experimental
/product="alpha-galactosidase"
/protein_id="CAA83665.1"
/db_xref="GI:475965"
/db_xref="SWISS-PROT:P43467"
/translation="MSLITVDQANRVFHLHNOTLSYIFAVEQGGTSLHLYFGGHVDHY
HGLRYPVRDGRFSGNLPGSTDRTFSDRTLPRKEYSTAGEMDYHLPAAIIVRHTDGANAL
VLVYQYRIEACKPKLSGLPAAFVEDETEAETLTIVLDQVSOVEDLOYTYIYRDPV
VRSVOVCQGHVHNLKXVASQIDFTDQRETYIILPCAHANERHPERGSINYGIQT
FSLRGTSSGHMFFLAVDHTTFESGDAIYFNLVSGNHAFELKEDQLDQLHLVWG
INSYNNWOLKAGATFQTPPEVLMVYTNKGLNAMSQAYHHLIRERVVRSEFKQERIV
VNWEATFFDNEAKLPIVDEAKQLGIEFMFLDDGFWGDRDDNSLSLGDWQVDRKF
PQCLNHFVYVHEQGLKFGIWLPEPMSYSDSKLYOOHPDYLMQVPGRSPSRNRYIL
DLGROANVRINFDQLDLKSKOIDYIKWDMNRHLSDIYSVALPERQGEVYRVYVIG
LYELLERLTAVPHILFEGSGGGFRDAGMAYYMPQIWAISDNTDAVRLTYQYTSL
APISLATHVSVSPNQQTRETSNSTRSAVASGVGLYELDLTQLSSADQKQVQKQV
VQYKQIRLPIQGEFFLRKSPITSNOQAWMFVSQODEAIVMVFNLTSYVAPSLNTK
LVGLNKLNPQNIATKAIFGDELQMGFYDPVYQDYTKVYHFKAVTEN"
terminator
8265..8289
/standard_name="terminator 3"
/citation=[1]
/notes="DR2"
/repeat_unit
9403..9426
/notes="DR2"
/citation=[1]
/pt_type=DIRECT
/complement(9519..10424)
/genes="iso-IS30"
/complement(9519..10424)
/partial
/genes="iso-IS30"
/standard_name="IS-element"
/function="IS-30 (E.coli) like IS element"
/citation=[1]
/codon_start=1
/transl_table=11
/product="insertion element"
/protein_id="CAA83666.1"
/db_xref="GI:475966"
/db_xref="SPTREMBL:Q52261"
/translation="SDAVTKRSASLKWCLPOTHLGLLDLLSSITYSERIKIETFCEL
GLSNQMVRLNRSPSTIYELSRQPIYQAEIAQDAEYKRSQCKRKTLSDELKQKI
LNLHLSWSPMIAHEFKLATKSIYNLWNGRIGFSLNDLPEHVRQRNRPQDSRYN
QSLGRSIEQRPMFNORNRIGDFELDTVVGPRGHSKVALLLTIDRSKRLWAYRLKDR
TTATVNEALTKELTTFNGPVSHTVDGRGTEFSGLSVLESQYGIKTYICHAYTPADVVV
MNALIGYVVVFTLKGLVLSTLVKLI"
repeat_unit
10396..10442
/notes="DR1"
/citation=[1]

```

```

/repeat_unit
10397..10443
/notes="IR2"
/citation=[1]
/pt_type=INVERTED
10681..10716
/standard_name="terminator 2"
/citation=[1]
/complement(10727..11593)
/genes="scrK"
/complement(10727..11593)
/genes="scrK"
/standard_name="fructokinase"
/function="fructose 6-phosphotransferase"
/citation=[1]
/codon_start=1
/transl_table=11
/product="fructokinase"
/protein_id="CAA83667.1"
/db_xref="GI:475967"
/db_xref="SWISS-PROT:P43468"
/translation="MLIGAERGGTKFVCATGAENQVSDRISPIPTTPTVETWAVDD
YFTTHPVDAGIGSGFGPIGVNPHDPKXYITTPKPGWGDFFLGHLSQFNIPLYWT
TDVNEAAYGESMIGIAKDVPSNIYMTIGTGVGAGVISQNHIFNRPHTHELGHMLNRL
PGDDFKSNCPYHIDICLEGLAAGPAVGKRTGKAGKDIPVDDPVMPITDIYIAQACVNL
VAFAPDKIILNGVMNQRLFPMIREKFAAYLNGYEEVPLDDYIYPAGLGNNSGIAG
GULLAQAALKNA"
gene
13852..15989
/genes="scrA"
/complement(13852..15807)
/genes="scrA"
/standard_name="enzyme IIsr"
/function="enzyme II of the PTS system, sucrose specific"
/citation=[1]
/codon_start=1
/transl_table=11
/evidence=experimental
/product="enzyme IIsr"
/protein_id="CAA83668.1"
/db_xref="GI:475968"
/db_xref="SWISS-PROT:P43470"
/translation="MNHQVADRVLNAIKNNIQAHAACATRLRLVIKDESKIDQOAL
DDADVKGFTETNGOYIILIGPDVKNVDALITVGLKEVTPDDIKAVAAQGNKP
LMDLKVLSDFIPIVPAIVAGLLMALNVLTAHFLMAKSVVEYVYGLKGAEMIN
ANASAPFTFLPILLGFSATKREGNYPILGATWGMIMVLSLVNGYSVATMAKKNVY
HNVEGLHVAQAGYQGOVLPVLGVAFILATLEFFHFKHKAQDFDTTPMFAIVITGFL
TFTIVGLPVLRTVSLDALTNLGLVNLSTWIGMIGLISALVITGLHQTFFAIEIQL

```

```

Query Match      18.8%; Score 287.6; DB 1; Length 21839;
Best Local Similarity 58.8%; Pred. No. 8e-67;
Matches 539; Conservative 0; Mismatches 364; Indels 13; Gaps 2;

QY   4  atggcatctgcgcgttgctgcgttccttgccagtggttggttggttcaccgcaacaagcgt 63
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15324 ATGGCTAGTGCgcgtttactttcttgcgaatcctctTAGGATTTTCAGCAAGCGC 15265

QY   64  ttccgcggaatagttcctcgccgcgcgtattgtatgcgatggtgttcccgagcctt 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15264 TTCGTTGGCAACCCCTATCTGGGTGCTACG-ATGGGCATGATCATGCTTACATCACT 15206

QY   124  ggtgaacggtacagctggtgcgcgccaccatggtgcgcgcgaatgcgaatggtgctcct 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15205 AGTTAACGGCTATAGCTTGCACACGACCATGCGCGGCAAGATGTTTACTTGAACGT 15146

QY   184  gtttggttagatgttgccaaagccggttaccagggcacccgcttcttctgtggtggt 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15145 CTTTGGGTTACAGCTTGCACAAAGCGGCTATCAAGCGCAAGTGTCTCCAGTCTTAGGCGT 15086

QY   244  ttcttgattctgcgaacgatcgagaagttccttcgcacaagcgactcaagcgactgcaga 303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 15085 CGCCTTCATCTACGTACCCCTTGAAAAATCTTCCATAAACACATTAAGGGGCATTGA 15026

QY   304  ctctcgatactccagtgctgacgttgctgctcaccggatctcttacctatcatcgccat 363

```



```

/product="enzyme II sucrose protein"
/protein_id="CAB09690.1"
/db_xref="GI:4581477"
/translation="MNHQVAERILNAVGRDNIQGRHCAIRLRLVLKDTGVIDQBAL
DNDPDLKGFQFEAGQYQIIVPGDVNTVYEEF IKLGI SEASTADLUKEJAGSKQKNP
VMAVLKLLSDIFVPLIPALVAGLLMALNALTAEHLFATKSLVENFPMWKGFADIVN
TMSAEPFTMPILIGYSATKRFNGNPLGVAVVMVPLINGVNAEAIASHNTY
WDFGKFAQGVGOVLPVIGVAFILAKLRFHFKYLNDIDFTPTLLSVITGFL
TETIVGPAIRFVNSGLTDGLVGLYNTLGAJLMVFGGFYSIVVTLGHQSPFAIETML
ITNYQHSKGGDFIFPVACANQAQACATFALFVTKNKTALAPAGVSAIILGITE
PALFGNHLKIPFFIATGASALFSLFGLFHLVAVLSAGSLIGFISIKAGVNYQFM
ISIFSLFADVSDFVSKIMGKIAIKKKNQONATYOPEKVIIDPHVKSGLLAPI
NGVPIHIGDITVSMNGNFIQNVKGVKAGDGLLGSFDEKEIKKSGLDOTVIIVTN
SASYNEILPLSENVDIKVGEKILLN"
RBS      4838..4842
/gene="sack"
/4838..5720
/gene="sack"
/4848..5720
/gene="sack"
/BC_number="2.7.1.4"
/codon_start=1
/transl_table=11
/product="fructokinase"
/protein_id="CAB09691.1"
/db_xref="GI:4581478"
/translation="MSVYVGSIEAGTKFVLAIADHFHNIKKFKFATTPQETISK
IKYKKNRVSAGLGSFPIDLNLSKTYITSTPKVGNKINLVGQKEALDIPY
FTVDNAYAGEMKNTKIKNLVLTITGIGGAIONGYFIIIGIHSEHGHRINRHR
DVLTFEGICPFHGDCLLEGVAAGPSLEARTGILKEISDDPDWDLISYIAQAANAT
LTLAPRCIILGGVMEKPNKMSLIQKQFISMLNLIIDLPCSVKEYIRLPTVRENGSAT
LGNFYLAISLFTKE"
BASE COUNT 1858 a 915 c 1031 g 1996 t
ORIGIN

Query Match      16.4%; Score 250.8; DB 1; Length 5800;
Best Local Similarity 51.1%; Pred. No. 7.2e-57;
Matches 725; Conservative 0; Mismatches 647; Indels 46; Gaps 4;

QY  4  atggcactgcgcgttcgcttctgcaagtggttggtttccacgcgaacacgcgt 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3366  ATGAGTCGCCGCACCATTTACCTTTATGCCAAATTTTAAATAGGTATTCTGCCACAAAGA 3425

QY  64  ttccgcgcgaatgagttcctggcgccgcgtattgtagcgatggtgttcccgagctt 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3426  TTTGGCGGAATCCATATTAGGTGCTG-TTGTAGGTATGATTATGTTGATGCTGGACT 3484

QY  124  ggtgaacggtacgactggtgcgcacacatggtgcggtgcggaatggtgcct 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3485  TATTAATGGATATAATGTTGCTGAAGCAATCTCTAATCATACATGACTTATTGGGATAT 3544

QY  184  gtttggttagatgttgcacacgcggttacacaggccacgcttctctcgtgctggtg 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3545  TTTTGGTGTAAAGTTGGCCAGCGCTGGATATCAAGGACAAAGTCTCTCTCTAATGGAGT 3604

QY  244  ttcttgattctgcacacgatcagaagttctctgcacacgactcaaggcactgcaga 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3605  GGCTTTATCTTGTCTAACTTGAGAGATCTTTTCATATAATACCTTAACGATGCTATAGA 3664

QY  304  ctctcgtacactccagtgctgctgctgctcgcgcggtacattcctacattccatccat 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3665  CTTACACATTTACTCCGTTACTTTTCAGTCATCATAACTGGATTTCTCACATTCATATTGT 3724

QY  364  tggccacgaacatgcgtggtggtggtggtggtggtggtggtggtggtggtggtggtg 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3725  GGGACACGACACTTCGTTTGTATCAAAATGGCTTCAGACATGTTGTTGGTGGACTTTATAA 3784

QY  424  ttccggtggtccagtcggtggtggtggtggtggtggtggtggtggtggtggtggtg 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3785  CACATTAGGGGCACACTAGGAATGCTTGTGTTTGGAGGCTTCTACTCGGCAATAGTAGTAAC 3844

```

```

QY  484  tggctgcaccagtccttcccgcaaatgagctggagctgtttaaccag----- 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3845  TGAATTACATCAAGATTTTCGGCCATTGAACAATGTTAATCACAACTATCAACACAG 3904

QY  533  -----ggtggtatccttcctcgcacgcgctatctatgcttaataatcccccaggtgc 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3905  TGGTATTGGTGGAGACTTTATTTTCCAGTTCGGCCCTGCGCAATATATGCTCAAGCTGG 3964

QY  586  ggcattgttgagcagtggttcttctcgcgaagagtgaaagcctcaaggccttcgaggtgc 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  3965  TGAACCTTTTGCATTTTATTTGTTACTAAAAACATTAAGACAAAAGCTCTGCAGCTCC 4024

QY  646  ttccagtggtcctcgcgttcttctgattacgcgcgctgcgctcgttcggtgtaaaccttcg 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4025  AGCTGGTGATCTGCATTTCTAGGTATTACAGAACCTGCCTTATTTGGGATTATCTAAA 4084

QY  706  cctgcgctggcgcgttcttctcgcgttcggtacgcgcgctgcggtggtgccttcgattgc 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4085  ACTAAAATATCCGTTCTTTATTGCTCTGGGGCTTCAGCAATTTGTTCAATTTATATGGG 4144

QY  766  actcttaatatcaaggcagtgctggtggcgctgcaggttcttctggtgtgtttcttat 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4145  ATTATCCATGTCCTTGGGCTTAGTCTGGGATCCGAGGATTAATTTGGCTTTATCTCTAT 4204

QY  826  tgatgtccagatagtgatcttctggtggtgagttgttacccttcttcttcacatgcatt 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4205  TAAAGCTGGGTATAACTTACAATTTATGATTTTCGATATTTATTAGCTTTCTTATTCCTT 4264

QY  886  cgcgcgcagcagatgcttatgcttcttacttctggttcgcgcgcgcgcgcgcgcgcgcgc 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4265  TGTGTTTACCTCAATCTATGCTGGCGGATGGAAGCTAAATCTATTACGAAAGAAAAA 4324

QY  946  tgaacccgctgctccagtgctgcaggaacgcgaacgcgaacgcgaacgcgaacgcgcgcga 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4325  TAAACAGATGC-----AACACTCAATACCACTGAGAAAGTTATTAT 4369

QY  1006  atttcaaacgattccaccatcatccaggcacctttgaccggtgaagctattgcactgag 1065
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4370  CGATCCAGTTAAAGTGGCGAATCTCTGCTCCGATAAATGGATTTGTGATTTCTCTGTC 4429

QY  1066  cagcgtcagcagtgccatggttccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4430  TGATGTAAGTGAACCTGTTTCTCAAAGAAATTTATGGGAAAAAGGTATTGCAATCAAGCC 4489

QY  1126  aaccaaggcgcagtgatttctcgcgtgagtgaaagattggtggtcattcccatctgc 1185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4490  TAAATCTGGAGAACTTTCTCACACGCCGCGAGAAATTTATTATTGCTTATGAACAGG 4549

QY  1186  ccatgcttctgcagttcgcacacagcgtgaggtggttccaatgtgatatcttgatgca 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4550  TCACGCTTATGGTATAAAACAAAAAATGGAGCT-----GAAGTTCTTTTACA 4597

QY  1246  cattggttcgcacacagtaaacctcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4598  TATTGGAATAGATGACTGTTTCAATGAATGATAATGATGATGATGATGATGATGATGATG 4657

QY  1306  cgatgaagtcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4658  CCAGAAAGTAAACAGCAGGGGATTTACTAGGATCTTTTGATAAAGAAAGAAATCAAGAGAG 4717

QY  1366  aggttatgaggttaaccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  4718  TGGATTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4755

RESULT      8
LOCUS      AE004395/1 10264 bp    DNA          BCT          31-JUL-2000
DEFINITION  Vibrio cholerae chromosome II, section 52 of 93 of the complete
            chromosome.
ACCESSION  AE004395 AE003853
VERSION    AE004395.1 GI:9658068
KEYWORDS

```

SOURCE	Vibrio cholerae.	
ORGANISM	Vibrio cholerae	
REFERENCE	Bacteria: Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.	
AUTHORS	1 (bases 1 to 10264) Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Niernan, W.C. and White, O.	
TITLE	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae	
JOURNAL	Nature 406 (6795), 477-483 (2000)	
MEDLINE	20406833	
REFERENCE	2 (bases 1 to 10264)	
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Niernan, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES	Location/Qualifiers	
source	1..10264 /organism="Vibrio cholerae" /strain="N16961" /serotype="O1" /db_xref="taxon:666" /chromosome="II" /note="biotype: El Tor" complement(114..1553) /gene="VCA0653" complement(114..1553) /gene="VCA0653" /note="similar to GB:M76768 SP:P22825 PID:155262; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="PTS system, sucrose-specific IIBC component" /protein_id="AAF96554.1" /db_xref="GI:9658069" /translation="WDYPVIAKQLLESIGGKSNIOALAHCASTRRLRLVNDFTQINESA IESLGVGQFQKVAQGYQIIFGSGIVNQVYAEAKLTGLVEMSTNDVASAGAEKQNW QRAVGLSDIFVIRIIPAVAGGLMLFNVLTPAGLFIPEGQSLIDANPGLADLASMIN TFANAPFVLPVLLAFSASRKFPGNPLGALMLVHPDLNGLGFGSASVSGVTPT WNILGFEIEKVGQSVLPVLSAYILAKIENGLRKIVPSVVDNLLTPLMAIFITGFL PTVVGPPLTRDVGMLGNLWYDPSAGFVGGLRGFIYAPVITGMHHSFATIEGFL LADIVTGTFTFEPTAAMSNIAQGAALAVGVTETKLKGVAIPSGVTALLGITEPA MGVNLKURYPFIAICGAALASAFITLFNRAQALGAAGLPGIISINPQQIGYYIMG MAISVPAFALTVLLAMREKTKQAQVTA" 1854..2846 /gene="VCA0654" 1854..2846 /gene="VCA0654" /note="similar to GB:L14685 PID:289864; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="sucrose operon repressor ScrR, putative" /protein_id="AAF96555.1" /db_xref="GI:9658070"	<p>/gene="VCA0655" /note="similar to GB:M74035 PID:155269; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="sucrose-6-phosphate dehydrogenase" /protein_id="AAF96556.1" /db_xref="GI:9658071" /translation="MLDTLLEAGGINNVTRILAPQGOVVLAKHPPLVPHLPDDVS LQSVLGEWQLSVORTAEVSDQOLAAGKAIEROKLETPLPYOTALDCYRPLWHISPP QGLLNDPFGFIYHQGEYHLFYQWHPFACHKDKYVWHLKSLDLVDMQWQSVALTIPSDW LDFHGVSFGHAVSHQDDLMWFYTGNTRLGVDRQRTMQCAARNMANGFEKILGPVIRC LPEGYMEHPDPELOGEAFFVPGOGIASANPHHTIEHONRIFRATQNAOQGIALLQG WPLDGSFDFYAPQATAGDRVLCGWMGLPDETQHPSCDQGWILHQLTALRELVREG TIYQPLRELDLQSEPHLLSDNVTELKTSFALQVLTLPWGCSRLRMQNTQYRVTL RUDANQLLRDRSATQIRQSDTIRELKDSTVELRLIADQSSLEIFINQGEHVMTS RIFTPDASGISLHGASVDKLYMAPASAPFNLEVNVP" complement(4690..5661) /gene="VCA0656" complement(4690..5661) /gene="VCA0656" /note="similar to GB:M76768 SP:P22824 PID:155263; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="fructokinase" /protein_id="AAF96557.1" /db_xref="GI:9658072" /translation="MNALVRLSSNHLRSDMSRVMLTGDVVDLTPDQQHLYLKC GAPANVAVALRCLGRSAFFRGVNDPFGREMQOQLTDRQVDCQHLYPDPVHRTSTV VDLDEHGERSTFMVKPSADQFLQISDIPSFONGEMLHVCSTALANQPSRSSTFAAIA QMKVGGYVSFDPNLRREVSWPEPQELQATVMRAVLADVVKESSEELQFLTQTSIEE GLQATADQIPLVVVTLGKAGALVVPNSQIVSGKAVKPIDTTGAGDAFVGGLLYRL SVAQDMHNQATILDVAVKWANGCGCALATTQKGAMTALPNOAALYAFLE" 6008..7546 /gene="VCA0657" 6008..7546 /gene="VCA0657" /note="similar to SP:P13035 GB:M21277 GB:M55989 GB:M96795 PID:146181; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="aerobic glycerol-3-phosphate dehydrogenase" /protein_id="AAF96558.1" /db_xref="GI:9658073" /translation="MNAKSTRTQDEVLDLIIVGGINGAGIAADATGRLKVLGYDA KDFASATSSAKFLVGGGLRYLEHYEFLVSEALAEVLLKKAPHIVTPMRFLPHR PLRPAWMI RAGLFDNLGKRTSLPASHKVLNKGASVTKPEWQIGFEYSDCWDDAR LVILNMAQAEQGAELVNYCTVEKAERMGDLWHVTLDDQTOQRFERRSHALVNAAGP WVKQFLNENAHVSSPYGRLIOGSHIIVPRIHDEPQAYILQNEKRIEVEIPLYDDYS MIGTVDVEYKGPDKRVATDAERDYLIIISIVNKHFMREIARSDDIIAEFFSVRLCDD NSPQAITRDYTLSDQDAEAPLLSIFGKLTYYRKLGEAMKHLAPFPFKMAPWTA DAPLPGENFDYAALKNLQAAFPFITESVIERWLSYGRSTTOLLAGVTGIEDLGIA FSGELYQREIDYLCEKEFAHQAQDIFWRRSKLGNHDTSVVEVESYLQOKFHAQOPL KATV" complement(7648..9258) /gene="VCA0658" complement(7648..9258) /gene="VCA0658" /note="similar to GB:X61367 PID:48195; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="AAF96559.1" /db_xref="GI:9658074" /translation="MYTFALIKITLQQAEMSAYTPSAQOEVLVGDHDLVSTDLKG VITYCNDIFCRITAGFOADELLGNINVRHASKRPAKPAFADMWHHLKQGHANRGIYK TKSGGFYVDAYVTYQQGQLTGYOSVRAHAKRWEIATKAYQALAAKAGAKKIQ FKLHTSLRYALLGALMPALQAPQWQMASLLPAGVLVFRLOELVVRTPOOL KOWQNEYDSISRLIYSGADAFSVADYHKLMAKARTILGRMWDSPRLGELANQLHL TTQEVHQALAAQNSNIQAVTQATDAVESAAERVSSHTSAHQLIDQVQDCAETKHSI</p>


```
PAITOLISAFONGTGHGDFIVTASMANVAQGAATFAIYFLTKDKKMKGLSSSGVS
ALLGTEPALGVNLKVFPEFCALIGSASAAAGLLOVVAVLSGSAGFGFLSIIKA
SSIPYVVCLELISAFIAPVYGYGKYKAVDFRAEAAVEAIEVQPIPEASAAN
KAQVDEVLAAPLAGEAVELVSDVPVSSAMKGIKAIKPSGNTVYAPVDGTVOIAF
DTGHAYIKSDNGAEILIHIGIVSMEGKFEQKQADQIKRKGDLVLTGTFDSKIAE
AGLONTMFIYVNTADYASVETLASSGTAVAGDSLSLEVK*
2227..2252
misc_signal
BASE COUNT 740 a 401 c 513 g 854 t
ORIGIN
Query Match 15.1%; Score 230.8; DB 1; Length 2508;
Best Local Similarity 50.1%; Pred No. 1.9e-51;
Matches 779; Conservative 0; Mismatches 717; Indels 58; Gaps 6;
QY 2 teatggcatctgcgcgttcgcgttcttgcagtggttgcagtggttgcagcaacaagc 61
Db 704 TAATGTACAGAGCGCTTCTGCTTCTGCCCATTTTGGTGGGATTTTCAGCCGCTAAAC 763
QY 62 gtttcggcgcaaatgagttcctggcgccgcgcgctattggtatgctgagtggttcccgagc 121
Db 764 GTTTTGGTGCTAACCAATTCCTTAGGAGCTCA-ATCGGCATGATTATGTCGCACCGGG 822
QY 122 ttggtgaacggctacgagctggcgcca-----ccatggcgtgcg 160
Db 823 GCAGCAAAATATCATTTGTTTACAGCTAATGCTCCAAATTTCAAGGCTGCTACTATTGT 882
QY 161 ggcgaatgccaaatggttccctgttgggttagatgttgcgaagccggttaccaggc 220
Db 883 GCTTATACAGGATTTTGAATATTTTGGTGTGCTATCTACTCAGCAAGTTTATACTTAT 942
QY 221 accgtcttcctgctggtgttcttgattcttggaacgcatcgagaagttcctgcac 280
Db 943 CAAGTATTCAGTACTTGTGCTGTATGGCTTTTATCTATCTTAAAGAAATTTTCCAT 1002
QY 281 aagcgactcaaggcactgcagactcctgatcactcagtgctgacgttgctgcacc 340
Db 1003 AAGAGATGCCCTCAGCAGTTGACCTTACCTTTACACCTTTATGTCAGTTATTACT 1062
QY 341 ggattccttacattcatccatggcccgcaatgcgctggtggcgagtgctggca 400
Db 1063 GGATTTTTCACATTTATCTGTTTATGCTGCTGTTTGAAGAAAGTTTCTGACTGGCTTCA 1122
QY 401 cagcgctcacaggacttatgatatttcggtgtccagtcgcggtctgctcttcgctg 460
Db 1123 ATGGNAATCGTATGGCTGTATGATACAGCGGATTCCTAGGTATGGGTCTCTTTGGTGT 1182
QY 461 gtctactcaccatcgctacatcgctgctgcaccagtccttcccgccaattgagctggag 520
Db 1183 CTGTATTCACCTGTAGTTATGACTGTCTCCATCAAAAGTTTCCAGCTATTGAAACCCAA 1242
QY 521 ctggt-----taaccagggtgataccttcattcttcacagcgcacatc 562
Db 1243 CTCATTTACAGCTTTCCAAATGGTACTGTGCTATGCTGACTTTTATGTTTACAGCTTCA 1302
QY 563 atggcctaatacgccagggtcgcatgtttggcagtgattcttcctggcggaagtgaa 622
Db 1303 ATGGCCATGTGCGACAGCGCAGCAACCTTTGCANATTTACTTCTTAACAAAGATAAG 1362
QY 623 aagctcaagggtctgcagtgctgcagtgctcgcgtgttcttggtattacggagcct 682
Db 1363 AAGATGAAGGCTCTCATCTTCTTCAGGTGTGTCAGCGCTCTTGGTATTACAGACCT 1422
QY 683 gcgatcttcgggtgaaccttcgctgcgcttcttccttccttcggtatcggtaccgca 742
Db 1423 GCTCTTTTGGGGTGAATTTGAATATCCTTTCCCATCTTTTGTGCTCTGATGGCTCA 1482
QY 743 gctatcggtggcgcttgattgactcctttaataatacaaggcaagtcggtggcgctgca 802
Db 1483 GCTAGGCTGCGACCAATTCGGGGCTTACTCAAGTTGTAAGGTTGTCTCTGTTGGCGCA 1542
QY 803 ggtttcttgggtgtgtttcttatgatgctccagatagtgctgttcttgggtgca 862
```

```
Db 1543 GGTTCCTTGGTTTCCCTTCATTAAGGCAAGTCTTATTCATTTTATGCTGTGAA 1602
QY 863 gttgttaocttcttcctatgcagcttcgagcagtgattgcttacttacttggctgc 922
Db 1603 TTAATCAGTGTTCCTATAGCTTTTGCAGTAACCTATATGGCTATGCTAAGACGAAGCGTGT 1662
QY 923 cgcacggcgagcatgcatgcagatgcaaccgctgctccagtcgctgcaggaacgacaaa 982
Db 1663 GATGCTTTTGGCTGCAAGCTGCTGCGAAGAGCTATTGAAGAGTGCAAGAAATTCGA 1722
QY 983 gccgaagcagaagcaccgcagaaattttcaaacgattcca---ccatcatccaggcacct 1039
Db 1723 GAAGAAGCTGCTTTCGACCAATAAAGGCACAAAGTTACTGATGAAGTCTTTCGGGCTCT 1782
QY 1040 ttgaccggtgaagctattgcactgcagcgcgtcagcgcgtatgcttcgttgcgcggaag 1099
Db 1783 CTGTGCTGGTGAAGCTGTGAATTAACCTCTGTTAATAGCCCTGTTTCTAGCGAAGCA 1842
QY 1100 ctggtcgcggcggttgcctatgcctcccaacaaaggcgaggttagtttctccgctgagtgga 1159
Db 1843 ATGGGTAAAGGGATTGCTATCAAACTAGTGTATATACAGTTTATGCACCGGTGTATGGG 1902
QY 1160 aagattgtggtgcctcccatctgcctgcttcgcagttcgcacccaaggctcaggat 1219
Db 1903 ACTGTTCAAAATGCTTTTGATACTGCGCATGCTTATGGTATTAAAGTCACACAATG---- 1957
QY 1220 gttccaatgtgatatcttgcacattggttgcacagtgtaaacctcaacggcacg 1279
Db 1958 -----GTGCGGAGATTCTTATTATATGTTGTTATGATACAGTATCAATGGAAGTAA 2010
QY 1280 cactttaaccgctgaagaagcaggcgcgtgaagtcacaaagcagggagcgtcgtgtgaa 1339
Db 2011 GGATTTGAACAAAAAGTTCAAGCAGATCAAAAAATTAAGAAAGGTGACGCTTCTCGGAACA 2070
QY 1340 ttcgatattgctgcttaaggctgaggtatgaggttaaccacgcgctggttcttcg 1399
Db 2071 TTTGATAGTGACAAATTCGAGAAGCTGCTGTGATAATAACAACAATGTTTCTTGTACT 2130
QY 1400 aattacaagaacacggacctgttaaacattacggttcttggcgaaattgaaagcgagcc 1459
Db 2131 AATACAGCTGATTATGCGTCAGTTGAACCCCTCGCTTCTTTCAGGAACCTGTTGCTGTAGGT 2190
QY 1460 ---aacctgctcaacgctcgaagaagcagcgggtgctccagcaacacccaagct 1510
Db 2191 GATAGCTTACTGAAGTTAAAAAATAAGAAATATTATCAGAAAGACCGTAAGGT 2244
RESULT 10
AE006222/c
LOCUS Pasteurella multocida PM70 section 189 of 204 of the complete genome.
DEFINITION Pasteurella multocida.
ACCESSION AE006222 AE004439
VERSION AE006222.1 GI:12722266
KEYWORDS
SOURCE Pasteurella multocida.
ORGANISM Pasteurella multocida.
REFERENCE 1 (bases 1 to 10085)
AUTHORS May, B.J., Zhang, Q., Li, L.L., Paustian, M.L., Whittam, T.S. and Kapur, V.
TITLE Complete genomic sequence of Pasteurella multocida, Pm70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 21145866
REFERENCE 2 (bases 1 to 10085)
AUTHORS Zhang, Q. and Kapur, V.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
FEATURES
Location/Qualifiers
```

source	1. .10085 /organism="Pasteurella multocida" /strain="PM70" /db_xref="taxon:747" 344. .625 /gene="PM1841" 344. .625 /gene="PM1841" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03925.1" /db_xref="GI:12722267" /translation="MNCRIGSLLEILHKGHQSIVNCRIGSLEKDSRAGRSYVNC RIGLSLACIAFNFSTYVNCRIGSLEKQNLHLFHLVDVNCRIGSLERIY" 1158. .1400 /gene="PM1842" 1158. .1400 /gene="PM1842" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03926.1" /db_xref="GI:12722268" /translation="MOCASFNTDRVNCRIGSLSEIFSQDCLLPDVCNCRIGSLENOKNC IAKFVNCRIGSLENNVFRNATFVNCRIGSLEM" complement(3259. .3224) /gene="PM1843" complement(3259. .3224) /gene="PM1843" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03927.1" /db_xref="GI:12722269" /translation="MKICLVKTSMSGDILHSLPALTDQARALPNLQVDWVVEESFAEI PWHSAVRVITIALRWRKPTPTALTTRKOWTYRTLQSEQYDAVIDAQLLKSALE ATRLARIGTHDRHSIREPIAACFYDKKAIYSQOHAVERIQLFAQSLNPLPHTQ GYGIARHPQHAQTPYVLFHGTTRQDKFLPDQOWRTLAEKLTAGQIQIHVPWSNS QEKQARIAGQLDLVILPKSTLSELAQIYNASAVSVDTGLAHLASAALDKTNLT YGATDKLIGSGKNGHYLVAESMDKIEADEIFQKLTALLSLNTI" complement(3221. .4267) /gene="rfaF" /note="PM1844" complement(3221. .4267) /gene="rfaF" /codon_start=1 /transl_table=11 /product="rfaF" /protein_id="AAK03928.1" /db_xref="GI:12722270" /translation="WHILIIGPSWVCDMMHSLYOTLKQHYPLCEIDVMAPNCKPL LRMPVROAITWPLGHGAFALGDRYRGKALRHQYDMAIVLPNSLKSAFIEFFAKIP LRGKWGESRYFLINDLRNKKDIPMMVQRYIALAFEGKAVPAEDLPRPVLYTDP VTOOHTLKFEKOTALISPRSGFCPGAEPGARWPHYHAKLAEMLIQOQYAIRL FGSAKDEAVGEIRQALPETMRHFCINLAGQTSLENAVSLIADCTAIVNDSGLMHIA AAVQRLVALYGTPTPTPLSEKAVIIRIEGGLIKVRKCKDSDEGYHQSLIDIQ ELVFQKLTGLANT" complement(433. .4743) /gene="PM1845" complement(433. .4743) /gene="PM1845" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03929.1" /db_xref="GI:12722271" /translation="MKSCTQLLFILTALPTSLFAIHVEQAFMSAKAGEPSAIFMN IHNTQNVDSLVQSDIRADIVLHGTOHGKMIETVGINLPAHQMTALKRGLHWVF NVEQDLNVGDSFPLRLFDNGEIMHEAKIIQY" complement(4859. .6283) /gene="ptsB" /note="PM1846"	CDS	complement(4859. .6283) /gene="ptsB" /codon_start=1 /transl_table=11 /product="ptsB" /protein_id="AAK03930.1" /db_xref="GI:12722272" /translation="MNFSQLAQVQIDKLGKDNISAAAHCASTRLRIVVKNENLIDKKG IENIEGVKGFQAVAGQYQIFGSGTVNKVYAALSKLIGIDMTTSEVAAAGTEKQGLL ORLVKGLADIFVPIIPAIVAGLLMGIIHSMITAKGFFVEEKNVDLYPAIADLVDFIN TIANAPFVLPVLLGFSATRKGPNFFLGAALGNLLVHPALSDSGWNYALTAKGNIQY WHIFGLEIRGVYQGTIVPVLVASWVLATLEKNRNVPSFLDNLTLPFLFALFITGLL AFTVPIGREAGSLTIGTLWYDTLGFVGAIGFTGLYAPITVITGMHQTFFAVETQL LAEVATGTGTFIPFAAMGNIQAAGAACLAAYVMKDVKVRGIAVPSGIIALLIGTEPA MFCVNLRYRYPFISAMIGAGISSVALENVKAIALGAAGLPGCIPSIKPDLSLAMYCVG MLISASIAFTLVILGKRAQLRAE" complement(6512. .7531) /gene="scrR" /note="PM1847" complement(6512. .7531) /gene="scrR" /codon_start=1 /transl_table=11 /product="scrR" /protein_id="AAK03931.1" /db_xref="GI:12722273" /translation="MSITNKKRITLSDIAKCCGLSTTTVSMILNNRGEFEKKAETCQ RVKATAEQOQYRANYAKSLKQWSNVIGFVLPDLTNYGFACHTAKILEKQCBENGLOL IACSDNNPQEKALIERLLDQIDFLITAPTHQDDPSYQKIGHTPTQLDRIYVPDL ALNYVISNDESHITALICRLDOYQLTEFMYGGQRTLSPSESRLAGFLQGLRQANLN LAKEWILHRTYQPSGYQMLAETAKLRLPQAIPTASTYLLGVLRYLLEQKIMDLL LNRQLHLATFDDHDILNALPFIHSTIKQNHLPQAHQIFQHIMRYFAGKKSPNQKIDCD VIYRD" complement(7557. .9056) /gene="scrB" /note="PM1848" complement(7557. .9056) /gene="scrB" /codon_start=1 /transl_table=11 /product="scrB" /protein_id="AAK03932.1" /db_xref="GI:12722274" /translation="MPFSLSKVRSFFCHFWTSPMYIFNOGKYKSLHAQOVGELAIQIRQ HYLRDTQRPRTYHLAPETGLNDPGLVDFGKHYHYFQWFPYAAIHGKHKHMFY DEHFQVADPLIPDELFEHSGCYSGALLWQDIIVAFYTGNTNENRQIRIPHQNIATF KADGILLKARCLLNAPTGYTEHRDPKPFITEEGKIRFILGAQREDLSGALTITYEMT DIDATPQLIGEEKLPDFDNQVFMMECPDLLKLQBDOLFWSPPQKAQKTYCFQNHYY ATYALGQYANLHFSQAHIIELDYDFDYAPQSIQSGEYIYAWGLPDITYPTDEYQW HSMLSLPROFSIQQGLIYQRPQVYAQCLPAHKYVVEGKQIQIIPHLDRSYVAFDAENQ PTLRFFTNQRGQQLILRYONGILTLDRTYTTQTALMQQFGQRHCILPOLNHFVEFF DRSVVEIFLNQSEKVTSPFFIADRENALDSSRLCVSAELKPIILTP" complement(9037. .9963) /gene="PM1849" complement(9037. .9963) /gene="PM1849" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03933.1" /db_xref="GI:12722275" /translation="MGKIWLGVLDVAVDLIPEGENTYLRCAAGAPANAVAGIARLATT GFIRGVDDPLGRFMLKQOENVCVTOYMLHPKORTSTVIYDLKEGERSFTFMVNP ADQFLELSDLPHPNDMLHCCSIALINEPSSSTFAAHRKAKSGGFSFDPNLR LWNSPAEMKTVYVNVVALADVLKFSHEELFLITDSDTLSDATAKLTQHYPKDLIIVTL GKNGATYIYQGSQVLVSKKALIPVDTTGADAFVSGLLFGISQVKNWONPALVEII KOANACALATTAKGAMSALPNKQDLDAFLAE" BASE COUNT 2923 a 1988 c 2212 g 2962 t ORIGIN	gene	gene	CDS	gene	CDS	gene	CDS	gene
--------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------	------	-----	------	-----	------	-----	------

Query Match 14.3%; Score 218.6; DB 1: Length 10085;
Best Local Similarity 54.5%; Pred. No. 4e-48;


```
QY 305 ttctgtatcactcgaagtctgtaagctgtctgctcgaacgattccctacattcgcgcatt 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1176 AACCTTTAAACGCCATTACTAGCGGATTTTCATAGCGGTTTGTAGCGTTTACTGTGTC 1235

QY 365 ggcacgaacaatgcgtggtggcgatgctgctggcacacagctctacagggaatttatgat 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1236 GGTCGGTTCACGGCTGATATCGGTTTCTATTTGGGTGATGCTCTGAACCTGGCTGTACAAC 1295

QY 425 ttcggtgtccagtcggtgctgctgctcttcggtctgctgctcactcaaccaatcgtcatcact 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1296 ACAGCTGGCTTGTGTGGTGGCGGTATTTGGCTTATTTATGCTCCGCTTGTATTAAT 1355

QY 485 ggtgtgcacagctcctccgcgaatgagctgagctggtt-----aacag 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1356 GGTATGATCACAGCTTCATTGGCATTTGAATCAAACTCAACTTCTTGGGATATTGCAACGACG 1415

QY 533 ggtggtatcctcattcgcgaacggaatcattggttaataatcgcgaacggtgcggaatgt 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1416 GCGGGCACCTTTATCTTCTTCTATTGGCGCGATGCTTAACGTTTCTCAAGGAGCGGCACG 1475

QY 593 ttggcaggttcttcctcggcggaagatgaaagctcaaggccttcagctgctcaggt 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1476 TTGGCGTGGCGGTGATGACCAAGATAAGAGATGAAGGATTTGGCATTCATTCCTCGT 1535

QY 653 gttcgcgtgttcttggtattacggagcctgcgatcttcggtgtaaccttcgcctgcgc 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1536 GTTACGGGTTTGGTGGGTATTACAGAACCAGCATGTTTGGTGTAACTTAAACACTACGT 1595

QY 713 tgcggttcttcacgtatcgttacgcgaatcagctatcggtagcgttggatgacctttt 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1596 TACCCATTTATTGCGGCAGTTTGGCTGCGCGGTATFCAAGTGGCTTCATCACAAGTTT 1655

QY 773 aatcaagcagcttgcgttggcgctgcaggttcttcgtggtgttcttattgatgct 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1656 AATGTGAACTCAGCATTTGGTGGAGCTGCTTCCAGCGATCATTTCCATCACCCTCA 1715

QY 833 ccagatatgctatgttcttgggtgtgagctgttacccttcttcacgattcgcgcga 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1716 GATAAGATTGGGTACTACATCGCGGTATGTTGATCGCGTCTCTTGACCGCATTTGTTCTC 1775

QY 893 gcatgcttatgg 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1776 ACCATTGTTCTTGG 1789

RESULT 12
SXSRA 2655 bp DNA BCT 10-FEB-1999
LOCUS S.xylosus scrA gene and unidentified open reading frames.
DEFINITION X69800
ACCESSION X69800
VERSION 1
KEYWORDS membrane protein; scrA gene; sucrose transport protein.
SOURCE Staphylococcus xylosus.
ORGANISM Staphylococcus xylosus.
REFERENCE 1 (bases 1 to 2655)
AUTHORS Wagner, E.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1992) E. Wagner, Mikrobielle Genetik,
Universitaet Tuebingen, Auf der Morgenstelle 28, 7400 Tuebingen 1,
FRG
2 (bases 1 to 2655)
AUTHORS Wagner, E., Goltz, F. and Bruckner, R.
TITLE Cloning and characterization of the scrA gene encoding the
sucrose-specific Enzyme II of the phosphotransferase system from
Staphylococcus xylosus
JOURNAL Mol. Gen. Genet. 241 (1-2), 33-41 (1993)
MEDLINE 94049686
FEATURES
    Location/Qualifiers
        1..2655
            /organism="Staphylococcus xylosus"
            /strain="DSM 20267"
```

```
/isolate="cured strain C2a"
/db_xref="taxon:1288"
/clone_lib="sub-genomic in Staphylococcus xylosus mutant
scr1"
/clone="pSE4"
complement(<1..121)
/note="ORF3"
/codon_start=1
/transl_table=11
/protein_id="CAA49459.1"
/db_xref="GI:407906"
/translation="SPTREMBL:Q56203"
/db_xref="MKSTTLCIHEFLNTHQRCIHQVILLFSLNKLKLDVTLNGY"
complement(132..137)
RBS 338..342
RBS 350..925
CDS
/note="ORF2"
/codon_start=1
/transl_table=11
/protein_id="CAA49460.1"
/db_xref="GI:407907"
/translation="MQHIYSEFYNQSKTKLKSVLGIVLHDDVNNNSAEDYIVWLQ
QRIDNNELEKGWACAYVDKQTCYFHPSPYVHWHGNAFANDHYIGTERCQSKINGIL
TDQFGENEESAYIAALLKKYQLPVNRETVKLHKMFEDTECPARAWSIHLDNAPT
EENIKMLQDYFIDKIKYHYFNKESDMFVG"
-35_signal 995..1000
-10_signal 1040..1045
RBS 1053..2495
gene
/genes="scrA"
CDS 1053..2495
/genes="scrA"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="EiIsr"
/protein_id="CAA49461.1"
/db_xref="GI:407908"
/db_xref="SMISS-PROT:P51184"
/translation="MNYKSAENILQALGEDNVEAMTHCATRLRLVLKDEGLVDEKA
LGDMDVVKGTGGGQVILIGSTVKNVSELEKITGKEASSVEKVTGKTNNMNF
ORFVKMLSDIFVPITPAIVAGGLIMINILITAPFYDNOSLIEYQNOFSLAEMIF
IFANAPFTLLPILIGFSAARFGGNAYLGAALGMILVHPLEMSAYDYPKALEAGKEIP
HWNLFGLINOVGQGVPLMVLATIEKLRKVIPTVLDNLITPLAILSTGF
ITFSFVGLTRTLGWLSDGLTWLYEFGAIGLIFGLLYAPIVITGMHHSFAIEQ
LADSSTGSGFPIPTATMSNINQGAALAAAFIIRKNNKLGKVASAGVSALLGITE
PAMFGVNLKIRYFFIGAIVSGSGISAYIAFFKVKALALGTAGIPGFTISISQNNQWLIH
YGIAMIIAIFAVFVTYALSRYKKYRNIEA"
BASE COUNT 909 a 347 c 521 g 878 t
ORIGIN

Query Match 13.8%; Score 210.6; DB 1; Length 2655;
Best Local Similarity 56.2%; Pred. No. 5.6e-46;
Matches 485; Conservative 0; Mismatches 359; Indels 19; Gaps 4;

QY 5 tggcatctgcgcgtctgcgttcttccagctgttggtaggtatggcgagcgtt 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1537 TTGCTAATGCCCATTTACATTTATTCCTATATACTAATTTGTTTGTGCTGCAAAACGAT 1596

QY 65 tcggcgcaatgagttcttcctggcgccgctattggtatggcgagctt 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1597 TTGAGGCAATGCATACCTCGAGCGGCAC-TTGGAATGATATTGGTTCATCCAGAAATTG 1655

QY 125 gtaacggtctacgactggtgcgcacacatggtcgcggcgaa---atgcaaatgtgtgttc 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1656 ATGAGTGCTTATGATTATCTCTAAAGCATTAAGAAGCGGGAAGAGATTCCTCATTTGGAAC 1715

QY 182 ctgttgggttagatgttgcgaacggttaccaggccggttaccaggccggttgcgtgtgtg 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1716 TTATTCGGTCTCGAAATTAATCAAGTAGGTAGGTACCAAGGTCAAGTCTTACCAATGCTAGTA 1775
```



```

/product="yfkf"
/protein_id="BAA22292.1"
/db_xref="GI:2443221"
/translation="MISYIVOTLIVCIAIYAYENKPNFRSANNLTQWAFSLLIAGSAFL
WIYMRVNPPLPRLGLHFKYIPF"
RBS
CDS
4563. 4579
4577. 5653
/codon_start=1
/transl_table=11
/product="yfkf"
/protein_id="BAA22293.1"
/db_xref="GI:2443222"
/translation="MDKTSAYOGFFGALYTLAYLGHAPILMTIESAKONAWHSYILG
VWIVIPALMLHRLMKHODKNIYELSDSSPIAGRIIILLESFYFLINAHDIRFFI
NLINILPRTPMAVLGGVIFVAICAREKETLRMAQIFLFPGLVLVLPFLPTLA
TQIELQNLTPVEGLIPYLOSYYAFGTMGELIILPLLSNRSPKLYTIFAILGAL
LLAVMFSSISVSPNLTSTFFDPAYVIRQIRITDLDRSDLIIAFAFWPIVIMVKA
GSLYIVVYGSLFKSLDKPAMYPPTGMSVVCWFPLNTNQLIDFNRIKPIINVI
SLLPLLIYLIKSKALFGAKAH"
CDS
complement(5650. 7065)
/codon_start=1
/transl_table=11
/product="yfla"
/protein_id="BAA22294.1"
/db_xref="GI:2443223"
/translation="MERLVWIEHISDMLGPPILILLTGLYFTILLKGFQRYPL
YIFQTTGSGVCKPKGEGTVPQLQALSSTIGAAINIVGPAAIMFGGGAFFWMW
LIALFAIRKESVSLAVHYREKNEQGEYVGPMPYITKGLRMKWLGVFFSVALIVEL
IPSIQVQNSVSLEATFFSPNKIYAGIGIAFLIGLVVIGVGRIGKVFEPVPLMAG
AYAGGLIIVLMNLSVPAPFSLVSNFASNAFTSSAVGGFAGAAEAETVWGFGRLYSN
EAGGTAPIAHAAMTDPHPROGFWISGIVIDTILICTTTAFIVLASGVWTKNASN
DPAALTTAFOHYGSGSGFYVSVLVEFVSTIMVIFYGVKQAEFLFGLAGHVIK
FYVLAIIIGAGGAKAIWGLVDLALFIVVPNVIALLLSRKKYALYTEFTSEQYY
LKDRTKQKPYPTPEAKNS"
RBS
terminator
CDS
complement(7063. 7077)
complement(7149. 7175)
complement(7190. 7450)
/codon_start=1
/transl_table=11
/product="yflb"
/protein_id="BAA22295.1"
/db_xref="GI:2443224"
/translation="MGYRIKKNAGMAEDGALKKELMTIDHLEIKCYDEKEREQLIRK
KSEQLLDKLVDRKGMFSKPSAFYKNKVYDRLGRNPSS"
CDS
complement(7432. 7584)
/codon_start=1
/transl_table=11
/product="yflc"
/protein_id="BAA22296.1"
/db_xref="GI:2443225"
/translation="MMDFSHIVSEDKIKRAIKDGFQNLPGMGKPLPKMTRHTCNRC
AWAIVF"
RBS
complement(7448. 7466)
RBS
complement(7582. 7596)
CDS
complement(7600. 7737)
/codon_start=1
/transl_table=11
/product="yfld"
/protein_id="BAA22297.1"
/db_xref="GI:2443226"
/translation="MREKVAKNAVESTFREDITCKTRYLSRNKGIKWYIENCMIKYK
V"
RBS
complement(7735. 7752)
RBS
7824. 7840
CDS
7838. 7987
/codon_start=1
/transl_table=11
/product="yflf"
/protein_id="BAA22298.1"
/db_xref="GI:2443227"
/translation="MKTFIKERGLAFFLIAVLLWIKTYVGVVLFNLFNGIDNTQKIL
LFVNPLSSSLFPLGFLFKKLOOTAIIVIFHLSFLLYANIVYRFPNDFITIPVI
MQAKTNGQLGDSAFSLMRPTDAFYFIDTIIILAIKVNKPAETSSSKSFRIFASS

```

```

ILVFLINLVAESRPELLTRSDRNVLVKVLYTNYFTIYDAVONIKSNORALADSS
DTEVENYKANYDPNVYFGKAEKKNVIVYSLESQSFIIDYKIDKEVTPFLNKL
AHDNETFFDNFFHOTGOGKTSDAEFMMNSLIPUAQGSVFNKAQNTLQSPALLKS
KNTYSATPHGNTQFTFNENMYKAGIDKDFDSDAYYDMEENTKNYMKDKPFFKESM
PLLESPLQPPFTYKFTLNSHFFPDAGDFGDSVVDNYFQSAHLDQSIHQ
FFNDLKKDGLYKAGSIIIVMGDHYGIGISENNKAMAKVLGKDEITLQAOALQVPLFIH
AAGVGEKVHKYAGDVAPITLHLLGVDTKDYILMSGSDILSKREHREIVPPRNGDIFIS
PKYTKISGKYYDTKTEKDESEVDKSEDSLVKKELEMSDKIINGDLRLFRYEPKGFKK
VNSDYDTTKHDEDSSETSKNEDK"
terminator
CDS
complement(9789. 9820)
complement(9820. 11178)
/codon_start=1
/transl_table=11
/product="yflg"
/protein_id="BAA22299.1"
/db_xref="GI:2443228"
/translation="MSEFLQKLGKSMFLPIAVLPAVGITILALGREDVFNIPFYQAGT
AVFHLPIFAIGAIGISKDSNGAAGLSAISLMDAAATKIDTNNMVAFGGIIA
GLIAGYTVNRFKDKLPEYLGFFSGGRRVLPILTAITITILAGIFGVWPPQSCINSF
GEMVLGGIGAGTGLFNRLIPLGLHVLNNIFWOFGEVNGVTGLARFFAKDPT
AGTYMGFFPIIMFGLPAACLAMVYVYAKPSKKATAGMMIGFALTAFTITGTEPIEFA
FMFLSPLLYAVHVLTVLSLFIWNWLGIRSGFSFAGAIIDYVLSYGAIEKPLULLLVG
ICYAAVEFIVFYVLKALNLTGPREDDVDVLDENTVDQVNNENIMLKGGLGKKNLQ
TIDHCATRLRLTKDTALVDEALKKAGAKGVKSGQSVQVIIGPNVEFAAEELRAA
VK"
RBS
complement(11176. 11195)
RBS
11334. 11352
CDS
11350. 12099
/codon_start=1
/transl_table=11
/product="yflg"
/protein_id="BAA22300.1"
/db_xref="GI:2443229"
/translation="MIVINDQELGKIGRIVALAREMKRKBEPGMSTKDLDLIG
AVLDEHGAVSAPEREYDPVTICISVNDVEAHGIPSTSKILKAGDLVIDISAEFEGF
YSDTGISFVLGGEERLHKQCAENAFQKGLQAKAGKQKQIGRAVYHEARSQGT
VIKLTGHGIGRSLHEAPNIMNYDPFDNALFKNGTVIALEPFIETRAETIVEAGDG
WTEKTPDKSMVAQVEHTIVTKDEPIILTKL"
RBS
12166. 12183
CDS
12181. 12495
/codon_start=1
/transl_table=11

```

Query Match 13.2%; Score 201.8; DB 1; Length 37900;
 Best Local Similarity 52.9%; Pred. No. 1.5e-43;
 Matches 482; Conservative 0; Mismatches 422; Indels 7; Gaps 2;

```

Qy 1 ctcattggcattcgcgcgttccttcgcagtgatgttggtggttcaccgcaacaag 60
  || || || || || || || || || || || || || || || || || || || ||
Db 985 CTGATTGCGAGGCACGCCCTTACGTTTTTTCCTGCGCTTAATCGATGGTCTGCCGTCAG 926

Qy 61 cgtttcggcggcaatgagttccttggcgccgctgatttggtatggcgatggtgtcccgag 120
  || || || || || || || || || || || || || || || || || || || ||
Db 925 CGGTTCCGGCGCAATCCGCTTCTCGGCATTGTGC-TCGGGGTTATGCTGCTGCATCTGA 867

Qy 121 cttggtgaacgctcagcagctggtgcgcaccattgctgcggcgaaatgccaatgggtc 180
  || || || || || || || || || || || || || || || || || || || ||
Db 866 TTTGCTGAATCGTGGGGATATGCGCGCGCAACAAAGAGAGATCCCGGTATGGAA 807

Qy 181 cctgtttggttttagattggtcccaagccggttacacggcgaccctgctcctgtgctggt 240
  || || || || || || || || || || || || || || || || || || || ||
Db 806 TCTATTGCGCTTAGAGTGCGAGAAAGTCGGCTATACAGGCCAGGTGCTCCCAATTTGCT 747

Qy 241 ggtttccttgattcttggaacgatogagaagttccttgcaacagcgactcaaggcactgc 300
  || || || || || || || || || || || || || || || || || || || ||
Db 746 TGCCTTCTTATGCTGGCGAAGATTGAGGTATTTTAAACAAAACGCACACCTGAGGGCAT 687

Qy 301 agacttcctgatcaactccagtgctgacgtgtgctgcacggattccttacattcctgc 360
  || || || || || || || || || || || || || || || || || || || ||
Db 686 ACAGCTGCTGCTGTGCGCACCATTACGCTTCTTTAACAGGATTTGCGCTTTTATCAT 627

Qy 361 cattggcccaagcaatgcgctggtggtgcacacggtgctgcacaggtctcacaggactta 420
  || || || || || || || || || || || || || || || || || || || ||

```

```
Db 626 TATCGGCCCGATTACATTTCGGATCGGTAAATGTAAGTACTGACTCAGGGCTTATCTCGGTGTT 567
QY 421 tgattcgggtggtccagtcggtggtctgctcttggctctggtctactacccaatcgtcat 480
Db 566 TGGTTCCTTTTGGCCGACTGGGCGTCTGTATATATGGCGGTTTCTACTCAGGCTCGTGAT 507
QY 481 cactggtctgcaccagtccttccgcgaattgagctgagact-----gttaaccagg 534
Db 506 TACCGGTATGCATCACACGTTTCTTGAGTCGACCTGCAGCTATCGGTTCAAAGCTCGG 447
QY 535 tgatccttcctcatctcgaacagcgcacatctatggctaatatgcgccagggtgcgcgatgtt 594
Db 446 CGGCACCTTTTATGGCCGATGCTGGCGCTTCTTAATATCGCACAGGTTGGCGGCTCT 387
QY 595 ggcagtgctcttcctggtgcgaagagtgaaagctcaaggccttgacaggtgcttcaggtgt 654
Db 386 GGCCATGATGTTTATCGTGAAGGATGAGAAACAAAAGGCCCTCTCCCTCACATCAGGGAT 327
QY 655 ctccgtgttcttggtattacgagcctgcgatcttctggtgtgaaccttcgcctgcgtg 714
Db 326 TTCTGCTTATCTTGGCATTACAGAGCCTGCCATCTTTGGGGTGAATCTGGGATACAGATT 267
QY 715 gccgttcttcacgtgtatcggtaccgcagctatcggtggtgcgcttggattgcactcttta 774
Db 266 CCCGTTTATCATTCGAATGTGTCAGCTCTGCGCTGGCGGAATGTATATTTCTTCTCAAGG 207
QY 775 tatcaagcgagttgctggtggcgctgcaggtttctcttggtgtgttctctattgatgtcc 834
Db 206 TGTCTGGCAAGCTCCGTCGGTGTGCGGCGGTGCCCGGGATTTTCTCAATCATGAGCCA 147
QY 835 agataggtcatgttcttgggtgtgtgcagttgttaccttcttcacgtcattcggcgagc 894
Db 146 GTACTGGGGCGCGTTTGCAATCGGATGGCTATTGTATTGATCGTGCCTTTGCCGGAAC 87
QY 895 gattgettatg 905
Db 86 ATACGCGTATG 76
```

Search completed: March 22, 2002, 08:49:30
Job time: 7679 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: March 21, 2002, 16:21:14 ; Search time 25.85 Seconds
(without alignments)
1379.098 Million cell updates/sec
Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MAMVFPVLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1058.5	44.8	651	2 S44257	phosphotransferase
2	953.5	40.4	664	1 B32243	phosphotransferase
3	701	29.7	480	2 S39978	scrA protein - Sta
4	695.5	29.4	630	2 H83686	PTS system, beta-g
5	673.5	28.5	636	2 C83724	PTS system, beta-g
6	673	28.5	632	2 S68599	phosphotransferase
7	659	27.9	479	2 F82432	PTS system, sucros
8	657	27.8	609	2 I40406	beta-glucoside per
9	646	27.3	479	2 JQ0781	sucrose uptake pro
10	645	27.3	609	2 T47097	hypothetical prote
11	640	27.1	631	2 B42603	beta-glucoside-spe
12	588.5	24.9	636	2 D86807	hypothetical prote
13	570	24.1	470	2 C69725	phosphotransferase
14	569.5	24.1	470	2 H83926	PTS system, trehal
15	567.5	24.0	625	2 C25937	phosphotransferase
16	489	20.7	456	2 S62331	phosphotransferase
17	487	20.6	458	2 H83881	PTS system, sucros
18	473	20.0	455	1 W0EBST	phosphotransferase
19	467	19.8	372	2 I39868	sac operon regulat
20	406	17.2	459	2 JU0293	levansucrase synth
21	390	16.5	460	2 A39938	phosphotransferase
22	367	15.5	473	2 C65236	phosphotransferase
23	360	15.2	473	2 A86122	PTS system enzyme
24	342.5	14.5	453	2 F84096	PTS system, sucros
25	341.5	14.5	692	1 S46953	phosphotransferase
26	339	14.3	478	2 B82263	PTS system, trehal
27	321.5	13.6	699	1 W0BSGS	phosphotransferase
28	317.5	13.4	651	1 S18607	phosphotransferase
29	314	13.3	195	2 D86766	hypothetical prote

30	305.5	12.9	675	2 D83755	PTS system, glucos
31	298.5	12.6	648	1 W0EC2N	phosphotransferase
32	297.5	12.6	648	2 E85567	hypothetical prote
33	289.5	12.3	675	1 S46952	phosphotransferase
34	287.5	12.2	631	2 D69750	phosphotransferase
35	285	12.1	521	2 C86678	hypothetical prote
36	283.5	12.0	446	2 B69744	sucrose phosphotra
37	281	11.9	454	2 C86766	hypothetical prote
38	271.5	11.5	173	2 C83839	PTS system, glucos
39	270.5	11.4	485	2 G65051	phosphotransferase
40	265.5	11.2	485	2 B85920	hypothetical prote
41	256	10.8	189	2 F70169	phosphotransferase
42	255	10.8	168	2 H69940	phosphotransferase
43	254.5	10.8	634	2 A32241	lactose transport
44	253.5	10.7	474	2 G85885	probable PTS enzym
45	250.5	10.6	474	2 D65017	hypothetical prote

RESULT

S44257

1

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pediococcus pen
C:Species: Pediococcus pentosaceus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999
C:Accession: S44257
R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.
A:Description: The sucrose and raffinose operons of Pediococcus pentosaceus PPE1.0.
A:Reference number: S44252
A:Accession: S44257
A:Molecule type: DNA
A:Residues: 1-651 <LEE>
A:Cross-references: EMBL:232771; NID:g493728; PIDN:CAA83668.1; PID:g475968
C:Genetics:
A:Gene: scrA
A:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera
C:Keywords: phosphotransferase
F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

ALIGNMENTS

C;Accession: F82432
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardonson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
 L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A;Accession: F82432
 A;Reference number: A82035; MUID:20406833
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-479 <HEI>
 A;Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN001
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A:Gene: VCA0653
 A:Map position: 2
 C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 27.9%; Score 659; DB 2; Length 479;
 Best Local Similarity 49.3%; Pred. No. 3.7e-38;
 Matches 133; Conservative 46; Mismatches 87; Indels 4; Gaps 1;

QY 1 MAMVPSLVNGDYVAATMAAGMPMWSLGLDVAAGYQGTLPVLVSVWILATIEKFLH 60
 Db 194 MLMVHPOLLNCGFGSASVSGTPTWNILGFEIEKVGYSGLPVLVSVILAKIENGLR 253
 QY 61 KRLKGTADFLITPVLTLTLTFTTIAIGPAMRWGDVLAHGLGLOGLYDFGPGVGLLFL 120
 Db 254 KIVPSVDNLTLMPLAIFITGELTFTVVGPLTRDVGMLNMLYDSAGFVGGLAFGF 313
 QY 121 VYSPVITGLHQSPFPPIELELF----NQGGSFIFATASMANIAOGAACLAFFFLAKSEKL 176
 Db 314 IYAPFVITGMHSHFIAETOLLADIVTGTFFPIAAMNSIAGAAALAVGVTKETKL 373
 QY 177 KGLAGAGSVAVLGITEPAIFGNLRLRWPFFIGTAAIGGALIALFNKIKAVAGAGF 236
 Db 374 KGVAIPSGVTALLGITEPAMFGVGNLRLRYPFIAAICGAALASAFITLFLNVKAQALGAAGL 433
 QY 237 LGVYSIDAPDMVFLVCAVVTFTFAAGAI 266
 Db 434 PGIISINPQGIYYIMGMAISFVAAPALT 463

RESULT 8
 I40406
 beta-glucoside permease - *Bacillus subtilis*
 C;Species: *Bacillus subtilis*
 C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 20-Jun-2000
 C;Accession: I40406; S65581; A69594; S47174
 R;Le Coq, D.; Lindner, C.; Kruger, S.; Steinmetz, M.; Stulke, J.
 J. Bacteriol. 177, 1527-1535, 1995
 A>Title: New beta-glucoside (bgl) genes in *Bacillus subtilis*: the bglP gene product has
 A;Reference number: I40406; MUID:95189730
 A;Accession: I40406
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-609 <RES>
 A;Cross-references: EMBL:X85408; NID:g1037169; PIDN:CAA59697.1; PID:g1037172
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1995
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beret
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emmerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrarri, E.
 Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
 A;Reference number: A69580; MUID:98044033
 A;Accession: A69594
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-287, 'F', 289-433, 'G', 435, 'S', 437-448, 'HR', 451-548, 'M', 550-551, 'S', 553-60
 A;Cross-references: GB:299124; GB:AL009126; NID:g2636442; PIDN:CAB15963.1; PID:g26364
 A;Experimental source: strain 168
 C;Genetics:
 A:Gene: bglP
 C;Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera
 F;458-609/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 27.8%; Score 657; DB 2; Length 609;
 Best Local Similarity 33.4%; Pred. No. 6.7e-38;
 Matches 153; Conservative 84; Mismatches 181; Indels 40; Gaps 8;

QY 13 DVAATMAAGMPMWSLGLDVAAGYQGTLPVLVSVWILATIEKFLHKLKGFADFLIT 72
 Db 186 DLTALLGAGK-PI-SFICLPVTAATYSTVTPILLSTWIASYVEKWDRETHASLKLIIV 243
 QY 73 PVTLTLTGTFTTIAIGPAMRWGDVLAHGLGLOGLYDFGPGVGLLGLVYSPVITGLHQ 132
 Db 244 PTFLLLVVPLTLITVGLGAILGEYLSSGVNVLFDHAGLVAMILLAGTSLIIMTGMHY 303
 QY 133 SFPIELELENQGG-SFIFATASMANIAOGAACLAFFFLAKSEKLGAGSAGVAVLGI 191
 Db 304 AFVPIMINNTAONGHDYLLPAMFLANNGAGASFAVFLRSRNNKFKSLATTSITLMGI 363
 QY 192 TEPAIFGVNRLRWPFFIGTAAIGGALIALFNKIKAVALGA-AGFIVGVSVSIDAPVMWF 250
 Db 364 TEPAMYGVMNRLKPKFAALIGGAAGAFYGMTGVASYIVGNGAGLPSIPVFIGPTPIYA 423
 QY 251 LVCAVVTFTFAAGAIYAGLYLVRNGSIDPDATAAPVPACTTKAEAEAPFNSDSTII 310
 Db 424 MIGLITAFAAETAAYLLGFEDVPSDGS-----QQPAVHEGSRRII 464
 QY 311 QAPLTGAIALSSVSDAMFASGKLGSGVAIVPTKQLVSPVSGKIVVAFSPGHAFVATK 370
 Db 465 HSPKGEVKALSEVKDGVFSAGVMGKGFATEPEGEVSVPRGVSITIFKTKHAIGITS- 523
 QY 371 AEDGSNVDIIMHIGFTDVNUNGTHFNPLKQKQDEVKAGELLCEDFIDAIKAAGYEVVTPI 430
 Db 524 ---DQGAELIHLIGLDTVKLEGQWFTAHIEKQKVAEPDPLVSPDLQIKAAAGYDVITPV 580
 QY 431 VVSNYKK--TGPVNTVGLGEIEAGANLLNVAKKEAVPA 466
 Db 581 IVTNDQYSFSPKEIG-----KVQPREALLA 607

RESULT 9
 JQ0781
 sucrose uptake protein - *Vibrio alginolyticus*
 N;Alternate names: enzyme II-sucrose protein
 C;Species: *Vibrio alginolyticus*
 C;Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 18-Jun-1999
 C;Accession: JQ0781
 R;Blatch, G.L.; Scholle, R.R.; Woods, D.R.
 Gene 95, 17-23, 1990
 A>Title: Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose uptake-
 A;Reference number: JQ0781; MUID:91071601
 A;Accession: JQ0781

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-479 <BLA>
A:Cross-references: GB:M76768; GB:M30194; NID:g155261; PIDN:AAA27555.1; PID:g155262
C:Genetics:
A:Gene: scrA
C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 27.3%; Score 646; DB 2; Length 479;
Best Local Similarity 46.5%; Pred. No. 2.9e-37;
Matches 128; Conservative 54; Mismatches 87; Indels 6; Gaps 3;

QY 1 MAWPELVNGYDVAAATMAAGEMPWLSFLGLDVAQAGYQGTVPVLVYVSWILATIEKFLH 60
DB 194 MLWVHPDLLNGWFGGASVSGNIPWNILGFEIQKVGQSVLPVLVSAFILAELVELGLR 253
QY 61 KRLKGTADPLTPVLTLLTGLTFTATGAPAMRWGDVLAHGLQGLYDFGPGVGLLFL 120
DB 254 KVIPSLDNLTPLLAIFTAGLLTFTVVPFTRDIFGLDGLNWLNTAGFVGAVFGL 313
QY 121 VYSPVITGLHOSFPPIELELF-----NOGGSFIFATASMANIAQAACLAIVFLAKSEKL 176
DB 314 IYAPFVITGMHSHFIAETQQLADIATTGCTEIFFPLAAMSNVSOQAALAVAGVMSKDKKM 373
QY 177 KCLAGASGVAVLGIETEPATGCVNLRWRPFIFIGICTAIGGALIALFNKAKAALGAAGF 236
DB 374 KGIAIPSGVTGLGITPEAPMGVNLKLRYPFFIAACAAALSSAFITMFNVKQAALGAAGL 433
QY 237 LGVWSIDAPDMVNLVCA-VVTFEFAAGAAIAYGL 270
DB 434 PGLIISI-TPDKIGYIAGVIAFLTAFLTVILGI 467

RESULT 10

T47097
hypothetical protein N17C [imported] - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T47097
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contain
A:Reference number: 224350; MUID:95219088
A:Accession: T47097
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-609 <YOS>
A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06652.1; PID:g603778
A:Experimental source: strain BGSC1A1
C:Genetics:
A:Gene: N17C
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 27.3%; Score 645; DB 2; Length 609;
Best Local Similarity 33.6%; Pred. No. 4.5e-37;
Matches 154; Conservative 86; Mismatches 178; Indels 40; Gaps 11;

QY 13 DVAATMAAGEMPWLSFLGLDVAQAGYQGTVPVLVYVSWILATIEKFLHRLKGTADFLIT 72
DB 186 DLTALLGAGK-PI-SFGLPVTAAVYSSVTFIPILLISWIASVERWIDRTHASLKLIVV 243
QY 73 PVLTLLLTGLTFTATGAPAMRWGDVLAHGLQGLYDFGPGVGLLFLVYSPVITGLHQ 132
DB 244 PFTLLIVVPLTLITVGLGAILGEYLSGVNLPDHAGLVAMIFLAGTFLSLIIMGMY 303
QY 133 SPPIELELFNQG-SFIFATASMANIAQAACLAIVFLAKSEKLKGLAGASGVAVLGI 191
DB 304 AFVPTMINNIAQNGHDIYLLPAMFMQAGASFAVFLSRNKKFKSLALTITSITAMGI 363
QY 192 TEPATGCVNLRWRPFIFIGTAAIGGALIALFNKAKAALGA-AGPLGVVSDAPDMVNF 250

DB 364 TEPAMYGVNMLKLPFAAALIGGAAGGAFYGMTGVASYIVGNAGLPSIPVFIGPTFIYA 423
QY 251 LVCVVTFPIAAGAAIAYGLYLVRNRGSDIDPATAPVPAGTTKAEAEAPAFESNDSTII 310
DB 424 MIGLIAF--AASTSAAYLL-----GPE-----VPS-----HRSQQPAYHEGSEII 464
QY 311 OAPLGEATALLSSVSDAMPASGLGSGVAIVPTKGLVSPVSGKIIVAPSPGHAFAPVTK 370
DB 465 HSPITGEVKALSEVDKGVFSAGVMKGFALPEEGEVSPVSGSVTTIFKTHAIGITS- 523
QY 371 AEDGSNVDILMHIGFTVNLNGTHFNPLKKQGVKAGELLCEFFDIDAIAKAGYEVTTPI 430
DB 524 ---DQGAELIHLGLTVKLEGWFTFAHMKESDKVAPGDPVLSFDLEQIKAKAGYDITPV 580
QY 431 VYSNKK--TGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
DB 581 IVTNTDQYSFSPVKEIG-----KVQPKALLA 607

RESULT 11

B42603
beta-glucoside-specific phosphotransferase system-dependent permease - Erwinia chrysae
C:Species: Erwinia chrysanthemi
C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 16-Jul-1999
C:Accession: B42603
R:el Hassouni, M.; Henrissat, B.; Chippaux, M.; Barras, F.
J. Bacteriol. 174, 765-777, 1992
A:Title: Nucleotide sequences of the arb genes, which control beta-glucoside utilization
olase family including enzymes from eubacteria, archaeobacteria, and humans.
A:Reference number: A42603; MUID:92121114
A:Accession: B42603
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-631 <ELI>
A:Cross-references: GB:M81772; NID:g148385; PIDN:AAA24814.1; PID:g148387
A:Note: sequence extracted from NCBI backbone (NCBI:77535, NCBIP:77537)
C:Genetics:
A:Gene: arbP
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase
F:479-631/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 27.1%; Score 640; DB 2; Length 631;
Best Local Similarity 33.8%; Pred. No. 1e-36;
Matches 154; Conservative 86; Mismatches 200; Indels 16; Gaps 8;

QY 3 MYFPSLVNGYDVAAATMAAGEMPWLSFLGLDVAQAGYQGTVPVLVYVSWILATIEKFLH 62
DB 184 LVHPSMIAAFN---AMQAPDHSTLHFLGIPITFINYSSSVIPILFASVWSCKLEKPLNRW 240
QY 63 LKGTADFLTPVLTLLTGLTFTATGAPAMRWGDVLAHGLQGLYDFGPGVGLLFLVY 122
DB 241 LHANIRNFTPLLCIVISVPLTFLLGPSATWLSQMLAGYQWLYGLNLSLLAGAVNGALW 300
QY 123 SPVITGLHOSFPPIELELFNQ--GGSFIFATASMANIAQAACLAIVFLAKSEKLKLAG 181
DB 301 QVCVIFGLHWGFVPLMLNFSVIGHDTLLPLVPAVLGGAGATGLVLLRTQDLKRGKIAG 360
QY 182 ASGVSAVLGTEPTEPPIFGVNLRLRWPFIFIGTAAIGGALIALFNKAKAALGAAGFLGVYS 241
DB 361 SAFAAIFGITEPAPVGVTLPLRRPFIFCGIGCALGAAMVGAHTMTYSGFPFSIFSFTQ 420
QY 242 IDAPDMVFLV-CAVVTFTAFG-AAIAYGLYLVRNRGSDIDPATAPVPAGTTKAEAE 299
DB 421 VIPPTGVDSVNAAVICTLLAFATAALTSWSFGVPKD---ETQPAADSPA--VLAETQA 475
QY 300 PAFESNDSTIIQAPLTGEAIALSSVSDAMPASGLGSGVAIVPTKGLVSPVSGKIIVAF 359
DB 476 NAGAVDETLF-SPLAGEVLLLEQVADRTFASGVMGKGAIRPTQGLRXPVDTVASUF 534
QY 360 PSCHAFVTRKAEKDGSDNLDLHGHGFTVNLNGTHFNPLKKQGVKAGELLCEFFDIDA 419
DB 535 KTHHAIGLASR----CGAEVLHVGIDTVLRDGRYTFPHRVGVGVVROGDLLEFDGPAI 590

RESULT 14

H83926 Pfs system, trehalose-specific enzyme II, BC component BH2216 [imported] - Bacillus halodurans

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000

C:Accession: H83926

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Y.

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and its complete genome sequence

A:Reference number: A83650; MUID:20263314

A:Accession: H83926

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-470 <STO>

A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA805935.1; GSPDB:GN000004

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2216

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 24.1%; Score 569.5; DB 2; Length 470;

Best Local Similarity 42.9%; Pred. No. 5, 4e-32;

Matches 112; Conservative 55; Mismatches 93; Indels 1; Gaps 1;

QY 1 MAWFPFSLVNGYDVAATMAAGEMPMSLFGDLVAQAGYGTVPVLVWVSWILATIEKFLH 60

Db 195 LMLVHPDLLNANGYGAQLEGEIPTWNLFGLEIQVYGGVQLPILFSAWILAKIEFLR 254

QY 61 KRLKGTADFLITPVLTLLTGLTFTAGPAMRWGDVLAHGLQGLYDFGPGVGLLFLG 120

Db 255 KRPDSIQLLVVAPVALLTGTIFAAGIGPTFTTGNGITNVTFSIFAAPVPLVGGFLYGL 314

QY 121 VYSPVITGLHOSFPPIELELFNQ--GGSFIFATASMANIAQAACLAFLVFLAKSEKLGL 179

Db 315 IYAPLVVGMHHTFLAVDLQLLIGTIGTFLWPLVLVLSNIAQSAALAMFAETDEKLGL 374

QY 180 AGASGVSALVGTTEPAIFGVNLRMPFFIGTAAIGGALIALFNKAVAGAGFLGV 239

Db 375 SLSSAVSALVGTTEPAMFGVNIYKFPFVCAIIAAGAFITVNGVLSIGVGLPGI 434

QY 240 VSDAPDMVMFLVCAVTFPI 260

Db 435 FSIQAGFWGVFFIGMVIAPIL 455

RESULT 15

C25977 phosphotransferase system enzyme II (EC 2.7.1.69), beta-glucoside-specific - Escherichia

C:Species: Escherichia coli

C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 18-Jun-1999

C:Accession: C25977; A47616; C65175

R:Schneetz, K.; Toloczky, C.; Rak, B.

J. Bacteriol. 169, 2579-2590, 1987

A:Title: Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide sequence, gene

s.

A:Reference number: A91833; MUID:87222180

A:Accession: C25977

A:Molecule type: DNA

A:Residues: 1-625 <SCH>

A:Cross-references: GB:M16487; GB:M60288; NID:g145414; PIDN:AAA23510.1; PID:g145418

J. Bramley, H.F.; Kornberg, H.L.

J. Gen. Microbiol. 133, 563-573, 1987

A:Title: Nucleotide sequence of bglC, the gene specifying enzymell(bgl) of the PEP:sugar

A:Reference number: A47616; MUID:88009877

A:Accession: A47616

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-625 <GBA>

A:Cross-references: GB:M15746; NID:g145412; PIDN:AAA83837.1; PID:g145413

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

Co

A.: Rose, D. J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: C65175

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-625 <BLAT>

A:Cross-references: GB:AE000449; GB:U00096; NID:g2367269; PIDN:AAC76745.1; PID:g17901

A:Experimental source: strain K-12, substrain MGI655

C:Genetics:

A:Gene: bglF; bglS

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

C:Keywords: phosphoprotein; phosphotransferase; transmembrane protein

F:473-625/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 24.0%; Score 567.5; DB 2; Length 625;

Best Local Similarity 30.8%; Pred. No. 1e-31;

Matches 142; Conservative 91; Mismatches 211; Indels 17; Gaps 7;

QY 2 AMVFPFSLVNGYDVAATMAAGEMPMSLFGDLVAQAGYGTVPVLVWVSWILATIEKFLHK 61

Db 180 ALVHPDLITAFENGQKADALGL---DFLGIPVTLNLSYSSVIPILFSAWILCSILERRNA 236

QY 62 RLKGTADFLITPVLTLLTGLTFTAGPAMRWGDVLAHGLQGLYDFGPGVGLLFLGLV 121

Db 237 WLPISAINKNFTFLCLLMVITPTVTLVGLPLSTWISLIAAGYLMVLYQAVPAFAGAVMGGF 296

QY 122 YSPVITGLHOSFPPIELELFN--QGSFIFATASMANIAQAACLAFLVFLAKSEKLGLA 180

Db 297 WQIFWFGHLHGLVPLCINNFTVLGVDYTMIPLLMPALMAQVGAALGVFLCERDAQKKVA 356

QY 181 GASGVSALVGTTEPAIFGVNLRMPFFIGTAAIGGALIALFNKAVAGAGFLGVV 240

Db 357 GSAALTSFGITEPAVYGVNLPKYPVFIACISGALGATIIGVYAQTKVYSFGLPSIFTEM 416

QY 241 SIDAPDMVMFLVCAVTFPIAFGAAATAGLYLVRRNGSIDPD-ATAAPVPAGTTKRAEAE 299

Db 417 QTPTSGIDFTVWASV-----IGGVIAIGCAFV---GTVMHFITAKROPAGAPQEKTP 468

QY 300 PAEFSNDSTIIIOAPLTGEAIALSSVSDFAMFASGKLGSGVAIVPTKGLVSPVSGKIIVAF 359

Db 469 EVITPPEQGIGCSPMTGEIVPLIHVADTTFASGLLGKGIALLPSVGEVRSVPAGRIASLF 528

QY 360 PSGHAFVATKAEKDSNVDILMHIGFTVNLNGTHFNPLKKGDEYKAGELLCFEDIDAI 419

Db 529 ATLHAIGI--ESDDG--VEILIHVGIDTVKLDGKFFSAHNVGDKVNTGDRLLISFDIPAI 584

QY 420 KAAGYEVTTPVVSNNYKKTGPVNTYGLGEIEAGANLLNVAK 460

Db 585 REAGFDLTTTPVLISNSDDFTDVLPHGTAQISAGEPLLSIIR 625

Search completed: March 21, 2002, 16:23:46

Job time: 152 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:21:59 ; Search time 29.76 seconds

(without alignments)
1062.320 Million cell updates/sec

Title: US-09-604-231-2

Perfect score: 2363

Sequence: 1 MAMVFFSLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 322656 seqs, 67552660 residues

Total number of hits satisfying chosen parameters: 322656

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_New.*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US05_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	660	27.9	627	1 PCT-US02-03987-13467	Sequence 13467, A
2	660	27.9	627	6 US-09-815-242-13467	Sequence 13467, A
3	660	27.9	627	7 US-10-072-851-13467	Sequence 13467, A
4	396	16.8	473	6 US-09-897-516-6783	Sequence 6783, Ap
5	345.5	14.6	679	1 PCT-US02-03987-5658	Sequence 5658, Ap
6	345.5	14.6	679	6 US-09-815-242-5658	Sequence 5658, Ap
7	345.5	14.6	679	7 US-10-072-851-5658	Sequence 5658, Ap
8	345.5	14.6	681	1 PCT-US02-03987-12270	Sequence 12270, A
9	345.5	14.6	681	6 US-09-815-242-12270	Sequence 12270, A
10	345.5	14.6	681	7 US-10-072-851-12270	Sequence 12270, A
11	338.5	14.3	484	1 PCT-US02-03987-10809	Sequence 10809, A
12	338.5	14.3	484	6 US-09-815-242-10809	Sequence 10809, A
13	338.5	14.3	484	7 US-10-072-851-10809	Sequence 10809, A
14	334	14.1	473	6 US-09-897-516-6540	Sequence 6540, Ap
15	323	13.7	484	1 PCT-US02-03987-12272	Sequence 12272, A
16	323	13.7	484	6 US-09-815-242-12272	Sequence 12272, A
17	323	13.7	484	7 US-10-072-851-12272	Sequence 12272, A
18	305.5	12.9	455	1 PCT-US02-03987-13794	Sequence 13794, A
19	305.5	12.9	455	6 US-09-815-242-13794	Sequence 13794, A
20	305.5	12.9	455	7 US-10-072-851-13794	Sequence 13794, A
21	300	12.7	687	1 PCT-US02-03987-5822	Sequence 5822, Ap
22	300	12.7	687	6 US-09-815-242-5822	Sequence 5822, Ap
23	300	12.7	687	7 US-10-072-851-5822	Sequence 5822, Ap
24	300	12.7	719	1 PCT-US02-03987-12942	Sequence 12942, A
25	300	12.7	719	6 US-09-815-242-12942	Sequence 12942, A

ALIGNMENTS

RESULT 1
PCT-US02-03987-13467
; Sequence 13467, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; TITLE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13467
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
PCT-US02-03987-13467

Query Match 27.9%; Score 660; DB 1; Length 627;

Best Local Similarity 34.3%; Pred. No. 8.1e-41;

Matches 159; Conservative 92; Mismatches 187; Indels 26; Gaps 10;

QY	1	MAMVFFSLVNGYDVAAATMAAGEMPMSLFGDVAQAQYQGTVPVLVSVILATIEKFLH	60
DB	182	MLLVSGSLFNAMAVA---QGGEVTAMNFFGF-IPVVGLOGSVLPFAFTIGVVGAKFEKAVR	237
QY	61	KRLKGTADFLITPVLTLTLTGLTFTIAGPAMRWGDLAHGLQGLYDFGGPVYGLLFLG	120
DB	238	KVVDVIDLLVTFPVILLVMSILGLFVGPFVHVENVILIATKAILSMFPGGLGFLIG	297
QY	121	VSPIVITGLHQSPPTIELELFNOGGSFTF-ATASMANIAQAACLAFFLAKSEKLKL	179
DB	298	VHQLVIVSGVHHFNLEVLQAAADHANPENAIITAAMTAQAATVAVGKTKPNKLT	357
QY	180	AGASGVSAVLGITEPAIFGVNLRWPFPGTIGTAIGGALIALFNKAVALGAAGFLG-	238
DB	358	APPAALSALFGLTEPAIFGVNLRWPFPGTIGTAIGGALIALFNKAVALGAAGFLG-	417
QY	239	VVSIDAPDMVFMFLVCAVVTFFIAFGAIAVGLVLRNRSIDDPATAAPVPACTKAAE	298
DB	418	MLTVGNGQFQYGLMVAVSFALGFALTYMEGY-----EDEVDATAAKQAEVAKKEE	470
QY	299	-APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGVAIVPTKGQILVSPVSGKIV	357
DB	471	VAPAAQLQNETLV--TPIVGDVVALADVNDPVFSSGANGQGIAPKPSGVVYALADAEVSI	528


```
QY 358 APPSGHAFVTRKAEDGNSVDILMHIGFDVTNLTGTHFNPLKQGDVEKAGELICEFDID 417
DB 529 APTGHAFLKTR----NGAEVLHVGIDVTSMNGDGFKAQGNKVKAGDVLGTGDSN 584

QY 418 AKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV 458
DB 585 KIAAGLDGDDTWMIVTNTADYASVAPVAT---GSVSRGDVIEV 625

RESULT 2
US-09-815-242-13467
; Sequence 13467, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13467
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13467
```

```
Query Match 27.9%; Score 660; DB 6; Length 627;
Best Local Similarity 34.3%; Pred. No. 8.1e-41;
Matches 159; Conservative 92; Mismatches 187; Indels 26; Gaps 10;

QY 1 MAVVFPVLNGYDYVAATMAAGEMPMSLFGDLVAQAGYQGTVPVLVVSILATIEKFLH 60
DB 182 MMLVSGSLPNAWAVA---QGGEVTAMNFFGF-IPVVGLOGSVLPFAFIIVGVGAKFEKAVR 237

QY 61 KRUKGTADFLITPVLTLLTGTFTTFAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFLG 120
DB 238 KVPVDVLDLVTPTVTLVMSILGLFVIGPVHVHVENYILIAATKAILSMFPGLGGFLIGG 297

QY 121 VYSPVITGLHQSFPPTELEFNQGGSFIF-ATASMANIAQGAACLAFFLAKSEKLKGL 179
DB 298 VHQLIVVSGVHHFNLNLEVLQVLAADHANPNENIITAMTAQGAATVAVGVKTNPKLKT 357

QY 180 AGASGSVAVLGIITEPAIFGVNLRWPFIFIGTAAIGGALIALFNKIKAVAGAAGFLG- 238
DB 358 AFPAALSAFLGITEPAIFGVNLRWPFIFLGLIAGIGGLASILGLAGTNGITTIPTG 417

QY 239 VVSDPADWVWFLVCVVTFVFFIAGAAIAYGLVLRNRSIDPDATAAPVPGTTKAAE 298
DB 418 MLXVNGQLPQYLLMVAVSPALGFALTYMFGY-----EDEVDTAAAKQAEVABEKEE 470
```

```
QY 299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIV 357
DB 471 VAPAAQLQNEVLV--TPIVGDVVALADVNDPVPSSGAMQGVIAVPVPSQGVYVALADAEVSI 528

QY 358 APPSGHAFVTRKAEDGNSVDILMHIGFDVTNLTGTHFNPLKQGDVEKAGELICEFDID 417
DB 529 APTGHAFLKTR----NGAEVLHVGIDVTSMNGDGFKAQGNKVKAGDVLGTGDSN 584

QY 418 AKAAGYEVTTPIVVSN---YKKTGPVNTYGLGEIEAGANLLNV 458
DB 585 KIAAGLDGDDTWMIVTNTADYASVAPVAT---GSVSRGDVIEV 625

RESULT 3
US-10-072-851-13467
; Sequence 13467, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE OF INVENTION: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13467
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-072-851-13467
```

```
Query Match 27.9%; Score 660; DB 7; Length 627;
Best Local Similarity 34.3%; Pred. No. 8.1e-41;
Matches 159; Conservative 92; Mismatches 187; Indels 26; Gaps 10;

QY 1 MAVVFPVLNGYDYVAATMAAGEMPMSLFGDLVAQAGYQGTVPVLVVSILATIEKFLH 60
DB 182 MMLVSGSLPNAWAVA---QGGEVTAMNFFGF-IPVVGLOGSVLPFAFIIVGVGAKFEKAVR 237

QY 61 KRUKGTADFLITPVLTLLTGTFTTFAIGPAMRWGVDVLAHGLQGLYDFGPGVGLLFLG 120
DB 238 KVPVDVLDLVTPTVTLVMSILGLFVIGPVHVHVENYILIAATKAILSMFPGLGGFLIGG 297

QY 121 VYSPVITGLHQSFPPTELEFNQGGSFIF-ATASMANIAQGAACLAFFLAKSEKLKGL 179
DB 298 VHQLIVVSGVHHFNLNLEVLQVLAADHANPNENIITAMTAQGAATVAVGVKTNPKLKT 357

QY 180 AGASGSVAVLGIITEPAIFGVNLRWPFIFIGTAAIGGALIALFNKIKAVAGAAGFLG- 238
DB 358 AFPAALSAFLGITEPAIFGVNLRWPFIFLGLIAGIGGLASILGLAGTNGITTIPTG 417

QY 239 VVSDPADWVWFLVCVVTFVFFIAGAAIAYGLVLRNRSIDPDATAAPVPGTTKAAE 298
DB 418 MLXVNGQLPQYLLMVAVSPALGFALTYMFGY-----EDEVDTAAAKQAEVABEKEE 470

QY 299 -APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIV 357
```

Db 471 VAPAAQNETLV--TPVGDVVALADVNDPVSSAMGGGIAVKSGQVVVALADAEVSI 528
Qy 358 APPSGHAFVTRKABDGNVDLMHIGFDTVNLNGTHFNPLKKGDEYKAGELICEFDID 417
Db 529 APPTGHAFGLKTR---NGAEVLIHVGIDTVSMNGDGFPEAKVQAQGNKVKAGDVLGTGFDSD 584
Qy 418 ATKAAGYEVTPPIVSN---YKKTGPVNTYGLGEIEAGANLNV 458
Db 585 KIAAGLDDTTMVTIVTNTADYASVAPVAT---GSVSKGDVIEV 625

RESULT 4
US-09-897-516-6783
; Sequence 6783, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6783
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6783

Query Match 16.8%; Score 396; DB 6; Length 473;
Best Local Similarity 31.5%; Pred. No. 1.6e-21;
Matches 88; Conservative 70; Mismatches 109; Indels 12; Gaps 5;
Qy 1 MAWVPSLVNGYDVNATMAAGBMP--WMSLFGDVAQAGVGGTVLPVLVVSVTLATIEKFL 59
Db 191 ITLVSPQLMNAVLQO-----QIPEWNEGFTTSKVGVOAQVIFSLKAGLGHIERL 245
Qy 60 HRLKGTADFLLTPVLTLLLTGFTTFAIGPAMRWGVDVLAHGLQGLYDFG-GPVGGLLF 118
Db 246 KKLVPDYLVLVVPVVSULLAVFLAHAIIGPIGRSIGDGVAMCVKGLMTGSPAPIGAALF 305
Qy 119 GLVYSPVITGLHQSPPIEELF--NOGGSFIFATASMANIAQGAACLAFFLAKSEKIL 177
Db 306 GFYAPLVITGVHQTALIDQMIQSTGGTTPWPIALSNTAQGSASVIGIHWASKQNER 365
Qy 178 GLAGASGVSAVLGITEPAIFGNLRLRPFETGIGTAAIGGALIALFNKAKVALGAAGFL 237
Db 366 ELSVPAAISAYLGVTEPAMYGINLKYRPMFCAMIGTSLAGLLCGLNHVTSNGIVGGIP 425
Qy 238 GVSTADPMVNFVCAVPTFFIARGAATAYGLYLVRN 276
Db 426 GILSIKQGFQWVIYLLAMLVITIV---PMALTVLMYRN 460

RESULT 5
PCT-US02-03987-5658
; Sequence 5658, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elitra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits Q
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02

; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
PCT-US02-03987-5658
Query Match 14.6%; Score 345.5; DB 1; Length 679;
Best Local Similarity 24.2%; Pred. No. 1.4e-17;
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;
Qy 27 SLFGDVAQAGYGGTVLPVLVVSVTLATIEKFLHRLKGTADFL-----ITPVLTLTLTG 82
Db 123 SILGIPTLQTGVFGGIIIGALAAW---CYNKFNINLPSYLGFFAGKREVPIM--NATTSF 178
Qy 83 LTFIAIGPAMRWGVDVLAHGLQ---GLYDFGGVGGLLFGLVYSPVITGLHQSF--- 134
Db 179 I--LAPPMALIW--PTIOTGLNAFSTGLLDSNTGVAVLFGFIKRLLPFGLUHFIHAPF 234
Qy 135 -----PPIELELFNQ-----GSFIFATASMANIAQGAACLAFFFLA 171
Db 235 WFEFGSWKNAAGEIITHGDORIFIEQIREGAHLTAGKFMQGEFPVMMFGLPAAALAIYHSA 294
Qy 172 KSEKLKLAGAGVSA-----VLGITEPAIFGNLRLRWPFFIGTAAIGGALIALFNK 227
Db 295 KPNKKVAVAGLSAALTSTLTGITEPLEFSFLVAPLILFFI---HAVLDGLSFLTLVLL 351
Qy 228 AVALG---AAGFLGVVSIDA-PDMVNF-----LVCVVVTFE 260
Db 352 DVHLGYTFSGGFDYVLGLVLPKNTQMWLVIPVGLVAVIYVYFVRFLVLVKKLYKTPGRE 411
Qy 261 -----AFGAATAYGLY-----LVR----- 274
Db 412 DKQSAVATASATELPYAVLEAMGKGNKIKHLDACITRLRVEYNDKSKVDVPGLKLGASS 471
Qy 275 -----RNGSIDPDATAAPVAGTTKAAEAEAPFSNDST 308
Db 472 VLEVGNNQAIQPKSDQIKHEMQQIMNGQVVENPTTMEDDKDETVAEDKSAATSELSH 531
Qy 309 ITQAPLTCEAIALSVSDAMFASGLKGVATVPTKQLVSPVSGKIVVAFPSGHAFAYR 368
Db 532 IVHADLTGVTPLSEVPDQVQVFSEKMMWGDIAIKPSGGEVRAFPNGKVMQVIFPTKHAIGL- 590
Qy 369 TRAEDEGSNDIILMHIGFDTVNLNGTHFNPLKKGDEYKAGELICEFDIDAIKAAGYEVTT 428
Db 591 ---VSDSGLELLIHIGLDTVKLNGEGFTLHVEGQEVKQGDLLINFDLDYIRNHAKSDIT 647
Qy 429 PIWVS 433
Db 648 PIIVT 652

RESULT 6
US-09-815-242-5658
; Sequence 5658, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5658

Query Match 14.6%; Score 345.5; DB 6; Length 679;
Best Local Similarity 24.2%; Pred. NO. 1.4e-17;
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;

QY 27 SLFGLDVAQAGYOGTVLPVLVSVWILATIEKFLHKKRLKGTADFL-----ITPVLTLTLTG 82
Db 123 SILGIPTLQGVGGIIIGALAAW---CYNKFYNINLPSYLGFGFACKRFVPIM-MATTSF 178
QY 83 LTFIATGPAMRWGVDVLAHGLQ----GLYDFGGPVGGLLGLVSPVITVGLHQSF---- 134
Db 179 I--LAPFMALIW--PIQTGLNAFSTGLDSNTGVAVLEFGFKRLLIPGLHHIFHAPF 234
QY 135 -----PPIELEFNQ-----GSFIFATASMANIAQGAACLAFFFLA 171
Db 235 WFEFGSKNAAGEIIRHGDQRIETIQIREGAHLTAGKFMQGEFFVMFGLPAAALAIYHSA 294
QY 172 KSEKLAGLAGASVSA---VLGITPAIFGVNLRWRWPEFFIGTAAIGGALIALFNK 227
Db 295 KPNKKVWAGLMGSAALTSFTLTGITEPLEFSFLVAPLLEFI---HAVLDGLSFLTYLL 351
QY 228 AVALG---AAGFLGVVSIDA-PDMVMF-----LVCVVVTFE----- 260
Db 352 DVHLGYTFSGGFDYVLLGLVLPNKTQWLVIIPVGLVAVIYVFFRELIVKLKYPGRE 411
QY 261 -----AFGAIAIYGLY-----LVR----- 274
Db 412 DKQSAVATASATELPYAVLEAMGGKANIKHLDACITRLRVEVNDKSKVDVPGLKDLGASG 471
QY 275 -----RNGSIDPDATAAPVPAGTTKAEAFAPAEFSNDST 308
Db 472 VLEGVNMQAIFGPKSDQIKHEMQQIMNGOVENPTTMDKDETIVVVAEDKSATSELH 531
QY 309 IIOAPLTGEAIALSSVDAMFASGKLGSGVAIVPTKQGLVSPVSGKIVVAFPSGHAFVR 368
Db 532 IVHAPLTGVTPLSEYDQVDFSEKMMGDGIAIKPSOGEVRAPENGKVMQIFPTKHAIGL- 590
QY 369 TKAEDGSNDVILMHIGFDVTNLTGTHFNPLKKGDEKAGELCEFDIDAIAAGVEYVT 428
Db 591 ---VSDSGLELLTHIGLDTVKLNGEGFTLHVEGQEVKQGDLLINFDLDYIRNHAESDIT 647
QY 429 PIVVS 433
Db 648 PIIVT 652

RESULT 7
US-10-072-851-5658
; Sequence 5658, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.

; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-072-851-5658

Query Match 14.6%; Score 345.5; DB 7; Length 679;
Best Local Similarity 24.2%; Pred. NO. 1.4e-17;
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;

QY 27 SLFGLDVAQAGYOGTVLPVLVSVWILATIEKFLHKKRLKGTADFL-----ITPVLTLTLTG 82
Db 123 SILGIPTLQGVGGIIIGALAAW---CYNKFYNINLPSYLGFGFACKRFVPIM-MATTSF 178
QY 83 LTFIATGPAMRWGVDVLAHGLQ----GLYDFGGPVGGLLGLVSPVITVGLHQSF---- 134
Db 179 I--LAPFMALIW--PIQTGLNAFSTGLDSNTGVAVLEFGFKRLLIPGLHHIFHAPF 234
QY 135 -----PPIELEFNQ-----GSFIFATASMANIAQGAACLAFFFLA 171
Db 235 WFEFGSKNAAGEIIRHGDQRIETIQIREGAHLTAGKFMQGEFFVMFGLPAAALAIYHSA 294
QY 172 KSEKLAGLAGASVSA---VLGITPAIFGVNLRWRWPEFFIGTAAIGGALIALFNK 227
Db 295 KPNKKVWAGLMGSAALTSFTLTGITEPLEFSFLVAPLLEFI---HAVLDGLSFLTYLL 351
QY 228 AVALG---AAGFLGVVSIDA-PDMVMF-----LVCVVVTFE----- 260
Db 352 DVHLGYTFSGGFDYVLLGLVLPNKTQWLVIIPVGLVAVIYVFFRELIVKLKYPGRE 411
QY 261 -----AFGAIAIYGLY-----LVR----- 274
Db 412 DKQSAVATASATELPYAVLEAMGGKANIKHLDACITRLRVEVNDKSKVDVPGLKDLGASG 471
QY 275 -----RNGSIDPDATAAPVPAGTTKAEAFAPAEFSNDST 308
Db 472 VLEGVNMQAIFGPKSDQIKHEMQQIMNGOVENPTTMDKDETIVVVAEDKSATSELH 531
QY 309 IIOAPLTGEAIALSSVDAMFASGKLGSGVAIVPTKQGLVSPVSGKIVVAFPSGHAFVR 368
Db 532 IVHAPLTGVTPLSEYDQVDFSEKMMGDGIAIKPSOGEVRAPENGKVMQIFPTKHAIGL- 590
QY 369 TKAEDGSNDVILMHIGFDVTNLTGTHFNPLKKGDEKAGELCEFDIDAIAAGVEYVT 428
Db 591 ---VSDSGLELLTHIGLDTVKLNGEGFTLHVEGQEVKQGDLLINFDLDYIRNHAESDIT 647
QY 429 PIVVS 433
Db 648 PIIVT 652

QY 429 PIVVS 433
Db 650 PIIVT 654

RESULT 10

US-10-072-851-12270
; Sequence 12270, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits
; FILE REFERENCE: ELITRA 028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12270
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-072-851-12270

Query Match 14.6%; Score 345.5; DB 7; Length 681;
Best Local Similarity 24.2%; Pred. No. 1.4e-17;
Matches 132; Conservative 68; Mismatches 192; Indels 153; Gaps 17;
QY 27 SLGLDVAAGYCGTGLPVLVSVILATIEKFLHKLKGTADFL---IPVLTLLLTGF 82
Db 125 SILGIPTLQGVFGIIIGALAAW---CYNKFYNINLPVLGFFAGRFVPIM-MATTSF 180
QY 83 LTFTAGPAMRWGVDVLAHLQ-----GLYDFGPGVGLLFLGVSPIVITGLHQSF--- 134
Db 181 I--LAFPMALI--PTIQTGINAFSTGLLDSNTGTVAVFLFGFKRLLIPTGLHHIFHAPF 236
QY 135 -----PPIELFLNQG-----GSFIFATASMANIAQAACLAFLFLLA 171
Db 237 WFEFGSKNAAGEIIGHDQRIIFQIREGAHLTAGKFMQGEFFVMFGLPAAALAIYHSA 296
QY 172 KSEKLKGLAGASGWSA---VLGTTEPAIFGVNLRRLRWPFFIGTAAIGGALIALFNK 227
Db 297 KPNKVVAGLMGSAALTSLTGTETPELPSFLFVAPLLEFFI---HAVLDGLSFLTYLL 353
QY 228 AVALG-----AAGFLGVWSIDA-PDMVMP-----LVCAVVTFFI----- 260
Db 354 DVHLGYTFSGGFIYVLLGLVLPNKTOWWLIPVGLVYAVIYVFFRELIVKLKYKTPGRE 413
QY 261 -----AFGAATAYGLY-----LVR----- 274
Db 414 DKSOAVTASATELPYAVLEAMGGKANIKHLDACITRLRVEVNDKSKVDVPGKLDLGASG 473
QY 275 -----RNGSIDPDATAAPVPAGTTRAEAPAEFSNDST 308
Db 474 VLEGVNMQAIFGPKSQDIKHMQQIINGVOENPTTMMEDDKDETIVVVAEDKSATSELSH 533
QY 309 IQAPLTGEATLSSVSAMDASFASKLGSVAIVPTKGQLVSPVSGKIVVAFPPSCHAFVR 368

Db 534 IVHAPLTGEVTPLEVPDQVPSERKMGDGIATKPSQSEVRAPNGKVMIPPTKHAICL- 592
QY 369 TKAEEDGNSVDILMHIGDFTVNLNTHENPLKKQDEYKAGSELLCEFDIDAKAAGYEVTT 428
Db 593 ---VSDSGLELLIHIGLDTVKLNGEFTLHVEEQEVKQGDLLINFOLDYIRNHAKSIT 649
QY 429 PIVVS 433
Db 650 PIIVT 654

RESULT 11

PCT-US02-03987-10809
; Sequence 10809, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits
; FILE REFERENCE: ELITRA 028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
PCT-US02-03987-10809

Query Match 14.3%; Score 338.5; DB 1; Length 484;
Best Local Similarity 33.2%; Pred. No. 2.9e-17;
Matches 94; Conservative 54; Mismatches 114; Indels 21; Gaps 10;
QY 2 AMYF---PSLVNGYDVAATMAAG---EMPWMSLFGDLVDAAGYOGTVLPVLVVSILATI 55
Db 201 ASVFGATPAL-GGVIGAVMTLTMGNPDAPISNIFTGTLTSLAG-OGGIIGVIFAVWLLSL 258
QY 56 EKFLHKLKGTADFLIIPVLTLLTGLTFTIAGPAMRWGVDVLAHLQGLYDF---GG 111
Db 259 EKOLHKVIPSIDIIVPTTISLVLGATIFLIMP---VAGATSNGLVGINVLEKGG 314
QY 112 PVGGLLEGLVSPITVITGLHQSPPPIELEFLNQG--SFIFATASMANIAQAACLAFLF 170
Db 315 MVAGFTLGLTFLPMVMEFLHQILTPHIEMINOTGMTLLPLILAMAGAGQVGAALALWIR 374
QY 171 AKSEK---LKGLAGASGVSAGVIGITEPAIFGVNLRRLRWPFFIGTAAIGGALIALF-NIK 227
Db 375 CKSDKKLVEMKALPV-GILGIGEPLIYGVTLPLGRPFTTACIGGIGGAVGAFGNVG 433
QY 228 AVALGAGFLGVWSIDA-PDMVMPFLVCAVVTFFIATFAGAAIAYGL 270
Db 434 AIAIGPSVALIPLIANNOWLAVVLGLLAAYAGGFVATLFFGI 476

RESULT 12

US-09-815-242-10809
; Sequence 10809, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2002-02-08

— — — — —

```

; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C
; REFERENCE: Bussey, Howard
; TITLE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08

```

Search completed: March 21, 2002, 16:27:36
Job time: 337 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:20:49 ; Search time 21.58 seconds
(without alignments)
488.023 Million cell updates/sec

Title: US-09-604-231-2
Perfect score: 2363
Sequence: 1 MANVFPVLNGYDVATMAA.....IERAGLLNVAKREAVPATP 468

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	113.5	4.8	1176 1	US-07-828-788A-12
2	113.5	4.8	1176 1	US-08-356-034-2
3	113.5	4.8	1176 1	US-08-933-891-2
4	113.5	4.8	1176 5	PCT-US92-11337-12
5	113.5	4.8	1179 1	US-08-040-751-1
6	108.5	4.6	731 2	US-08-911-364-1
7	108.5	4.6	733 4	US-08-464-700-2
8	108	4.6	1011 3	US-08-836-325-2
9	108	4.6	1989 3	US-08-836-325-12
10	107	4.5	1984 3	US-08-836-325-10
11	106.5	4.5	1129 6	5164180-6
12	106.5	4.5	1179 6	5188960-2
13	106.5	4.5	1864 2	US-08-804-227C-3
14	105.5	4.5	462 3	US-08-801-344-9
15	105.5	4.5	462 4	US-09-498-599-9
16	103.5	4.4	396 2	US-08-850-880-4
17	103.5	4.4	396 2	US-08-944-916-4
18	103.5	4.4	396 2	US-08-814-877-4
19	102	4.3	3033 1	US-07-925-695-9
20	101	4.3	3033 1	US-07-925-695-8
21	100.5	4.3	522 4	US-08-894-818B-3
22	100.5	4.3	525 1	US-08-356-340-2
23	100.5	4.3	525 2	US-08-786-555-2
24	100.5	4.3	654 4	US-08-894-818B-35
25	100.5	4.3	713 3	US-09-335-409-11
26	100	4.2	528 2	US-08-403-852D-21
27	100	4.2	528 3	US-08-510-646B-22

28 100 4.2 528 4 US-09-231-818-21 Sequence 21, Appl
29 100 4.2 1969 3 US-08-836-325-16 Sequence 16, Appl
30 99.5 4.2 792 2 US-08-678-039A-40 Sequence 40, Appl
31 99 4.2 756 4 US-08-960-048-10 Sequence 10, Appl
32 99 4.2 1275 3 US-09-120-513-2 Sequence 2, Appl
33 99 4.2 1275 4 US-09-450-105-2 Patent No. 5254799
34 98 4.1 1184 6 5254799-6 Patent No. 5254799
35 98 4.1 1188 6 5254799-7 Patent No. 5254799
36 97.5 4.1 2005 3 US-08-836-325-7 Sequence 7, Appl
37 96.5 4.1 593 6 5523211-1 Patent No. 5523211
38 96.5 4.1 648 3 US-08-810-720-13 Sequence 13, Appl
39 96.5 4.1 934 1 US-08-446-486-7 Sequence 7, Appl
40 96.5 4.1 934 1 US-08-463-308-7 Sequence 7, Appl
41 96.5 4.1 1176 1 US-08-446-486-6 Sequence 6, Appl
42 96.5 4.1 1176 1 US-08-157-363A-10 Sequence 10, Appl
43 96.5 4.1 1176 1 US-08-463-308-6 Sequence 6, Appl
44 96.5 4.1 1176 5 PCT-US93-11405A-10 Sequence 10, Appl
45 96.5 4.1 1176 5 PCT-US93-11405-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-07-828-788A-12
; Sequence 12, Application US/07828788A
; Patent No. 5273746
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: HICKLE, LESLIE A.
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/828,788A
; FILING DATE: 19920129
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA75
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1176 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: AIZAWAI
; INDIVIDUAL ISOLATE: PS811
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST STICK
; CLONE: 811A2
US-07-828-788A-12

```

; TELEFAX: (904)372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1176 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: AIZAWAI
; INDIVIDUAL ISOLATE: PS81I
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDAGEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 811A2
; US-08-356-034-2

Query Match 4.8%; Score 113.5; DB 1; Length 1176;
Best Local Similarity 22.4%; Pred. No. 0.019;
Matches 56; Conservative 43; Mismatches 96; Indels 55; Gaps 13;

QY 249 MFLVCVVVTFIAFGAAIAYGLYLVRNRSID-----PDATAAPVPAG-----TTK 294
Db 377 LFVLDTGTEFSASLTADLPSTIY--RQGTVDSDLVIPPDNSVPARAGFSHRLSHVTML 434
QY 295 AEA-----EAP-----A-EFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
Db 435 SQAAGAVYTLRAPTFSWRHSAEFSLIPSSQITQIPLT-KSINLGSCTSVVKGPGFTGG 493
QY 337 GVAIVPTKGQLVSPVSGKIVVAFPSGHAFVTRKAEDGSNDVILMHIGFDVNLNGHTFN 396
Db 494 DILRITSPGQISTL---RVITAPLSORYRIRY--ASTNLOFHTSIDGRPINQGNFS 548
QY 397 PLKKGQDEVKAGELLCFDDAIIKAAGYEVVTPVWSNYKKTGPVNTYGLGEIEAGANLL 456
Db 549 ATMSSGGLQSG-----SFTAGF--TTPF---NFSNGSSIFTLSAHVFNNG-NEV 593
QY 457 NVAKKEAVPA 466
Db 594 YIERIEFVPA 603

RESULT 3
US-08-933-891-2
; Sequence 2, Application US/08933891
; Patent No. 6096708
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Sick, August J.
; TITLE OF INVENTION: No. 6096708el Bacillus thuringiensis Isolate
; TITLE OF INVENTION: Active Against Lepidopteran Pests, and Genes Encoding
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/356,034
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; FILING DATE: US/08/210,110
; APPLICATION NUMBER: 07/865,168
; FILING DATE: 09-APR-92
; PRIORITY APPLICATION DATA:
; FILING DATE: 07/451,261
; APPLICATION NUMBER: 14-DEC-89
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; FILING DATE: 07/371,955
; APPLICATION NUMBER: 27-JUN-89
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, Roman
; REGISTRATION NUMBER: 21,023
; REFERENCE/DOCKET NUMBER: MA43.C1.D1
; TELEPHONE: (904)375-8100

```


GENERAL INFORMATION:
APPLICANT: Payne, Jewel M
APPLICANT: Sick, August J
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 54078
TITLE OF INVENTION: Lepidopteran-active Toxins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st STREET, SUITE A-1
CITY: GAINESVILLE
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/040,751
FILING DATE: 19930329
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REFERENCE/DOCKET NUMBER: MA39.C1.D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1179 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: AIZAWAI
INDIVIDUAL ISOLATE: PS81
IMMEDIATE SOURCE:
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
CLONE: 81R1
US-08-040-751-1

Query Match 4.8%; Score 113.5; DB 1; Length 1179;
Best Local Similarity 22.4%; Pred. No. 0.019;
Matches 56; Conservative 43; Mismatches 96; Indels 55; Gaps 13;
QY 249 MFLVCAVVTFFIAFGAIAIYGLYLVRNGSID-----PDATAAPVPAG-----TTK 294
Db 380 LFLVDGTEFSFASITADLPSTIY--RQRTVDSLDVIPPQDINSVPARAGFSHRLSHVTML 437
QY 295 AEA-----EAP-----AEFSN--DSTIIQAPLTGEATALLSVSDAMFASGKLGS 336
Db 438 SQAGAVYTLRAPTFSSWRHSABFSLNIPSSQITQIPLT-KSINLGSQTSVWKGPFTGG 496
QY 337 GVAIVPTKGQVSPVSGKIVVAPSPGHAFVTRKAEDGSNVDILMHIGFTVNLNGPHEN 396
Db 497 DILRTSPGQISLT---RVITAPLSQRYVRIRY--ASTTNLFQHTSIDGRPINQGNFS 551
QY 397 PLKQGDVAVKAGELLCEFDIDAKAAGYEVTTPIVVSNNYKKTGPVNTYGLGEIEAGANLL 456
Db 552 ATMSSGGLQSG-----SFRTAGF--TTPF---NFSNGSSIIFLSAHVFNESG-NEV 596
QY 457 NVAKKEAVPA 466
Db 597 YIERIEFVPA 606

RESULT 6
US-08-911-364-1
; Sequence 1, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,364
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,552
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-911-364-1

Query Match 4.6%; Score 108.5; DB 2; Length 731;
Best Local Similarity 20.6%; Pred. No. 0.03;
Matches 122; Conservative 47; Mismatches 227; Indels 195; Gaps 22;
QY 16 ATMAAGEMPWWSLPGLDVAAGYOGTVLPVLVSVNIIATIEKFLHKLKGTADFLLTPVL 75
Db 108 AGVKPGVPGVGLPGV-----YPGVLP-----GARFPVG----- 138
QY 76 TLLLTGLTFTATGAPAMRWYGDVLAHGLQGLYDFGPPVGGLLFGL-VYSPVITGLHQSF 134
Db 139 --VLPGVPTGAGVKPKAPGPGGAFA-GIPGVGPGPGVPLGYPKAPKLPGLPYGLPY 195
QY 135 PPIELEFNQGGSFIFATASMANIAQG-----AACLAVFFLAKSEKLGLAGA 182
Db 196 TTGKLP-YGYGPGVAGAGKAGYPTGTGVPQAAAAAATAKAAKFGAGAGVLPGVGGA 254
QY 183 SGVSAVLGITEPAIFGVNLRRLRWPFFIGTAAATGGALIALFNKAKVALGAAGFL----- 237
Db 255 -GVPVGPAGI-PGIGGI-----AGVGTAAAAAATAKAAKFGAGAGVLPGVG 301
QY 238 -----GVVSDIDPDMVWFL----- 251
Db 302 GFPGVGVPGAGVPGVPGVPGAGIPVVPVGGAGIPGAIVPVVSPAAAAAATAKAAKAGYARP 361
QY 252 ---VCAVVTFFIAFGAIAIYGLYLVRNG-----SIDPDATAAP-- 287

Db 362 GVGGGIPTYGVGAGGPGPGVGGVGGIPGVAGVPSVGGVPGVGVGGVGSPEAQAAAA 421
QY 288 -----VPAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVDAMFASG-KLGSGV 338
Db 422 KAAKYGVGTPAAAAAQAQAQGLVPGVGVAGVGVAGVGVAGVGVAGVGVAGVGVAGV 481
QY 339 AIVPTKGQLVSPVSGKIVVAFPSGHAFVTRKAEDSGSNVDILMHIGFDT-----VNL 390
Db 482 GVAPGVG--VAPGIG-----PGVAAAAAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 533
QY 391 NG-----THFNPLKKQGDVEVKAGEL-----LCEFD 415
Db 534 PGLGVAGVPGVGLGVAGVPGFAGAGDEGVRRSLSPELREGDPSSQHLPTSPSPRPGA 593
QY 416 IDAIIKAGYEVTTPIVVSNYKTKGTPVNTYGLGEIEAGANLLNVAKKEAVPA 466
Db 594 LAAAKAAYGAAPVGLGLGALGGVIGP-GVVGAGPAAAAAQAQAQAQAQAQAQAQAQAQA 643

RESULT 7
US-08-464-700-2
; Sequence 2, Application US/08464700
; Patent No. 6232458
; GENERAL INFORMATION:
; APPLICANT: WEISS, ANTHONY S
; APPLICANT: MARTIN, STEPHEN L
; TITLE OF INVENTION: SYNTHETIC POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,700
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PL6520
; FILING DATE: 22-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PL9661
; FILING DATE: 28-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU93/00655
; FILING DATE: 16-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GH3CUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-464-700-2

Query Match 4.6%; Score 108.5; DB 4; Length 733;
Best Local Similarity 20.6%; Pred. No. 0.03;
Matches 122; Conservative 47; Mismatches 227; Indels 195; Gaps 22;

QY 16 ATMAAGEMPWNSLEGLDVAQAGYQGTIVPLVIVSWILATIEKFLHKKLKTGADFLLTPVL 75
Db 110 AGVKPGKVPVGLPGV-----YPGGVL-----GARFPVGV----- 140
QY 76 TLLLTGFLTFTIATGPAMRWGDLVLAHGLGLOGLYDFGPGVGGLLFGL-VYSPITVITGLHQSF 134
Db 141 --VLPGVPTGAGVKPKAPGVGGGAFAGIPGVGPGGPGQVPLGPIKAKPLKPGYGLPY 197
QY 135 PPIELFLNOGGSFIFATASMANIAQG-----AACLAVFFLAKSEKILKGLAGA 182
Db 198 TTCKLP-YGYGPGVAGAACAGKAGVGTGTGVPQAAAAAQAQAQAQAQAQAQAQAQAQAQA 256
QY 183 SGVSAVLGITEPAIFGVNLRRLRWPFFIGITAAIGGALIALFNKIKAVALGAAGFL----- 237
Db 257 -GVPGVPGAI-PGIGGI-----AGVGTAAAAAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 303
QY 238 -----GVVSDIDAPDMWFL----- 251
Db 304 GFPGVVGVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPGVVSPAAAAAQAQAQAQAQAQAQAQA 363
QY 252 ----VCVVVTFEIAFGAAIAIAYGLVLRNG-----SIDPDATAAP-- 287
Db 364 GVGVGIGIPTYGVGAGGPGPGVGVGGVGGIPGVAGVPSVGGVPGVGVGGVGSPEAQAAAA 423
QY 288 -----VPAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVDAMFASG-KLGSGV 338
Db 424 KAAKYGVGTPAAAAAQAQAQGLVPGVGVAGVGVAGVGVAGVGVAGVGVAGVGVAGV 483
QY 339 AIVPTKGQLVSPVSGKIVVAFPSGHAFVTRKAEDSGSNVDILMHIGFDT-----VNL 390
Db 484 GVAPGVG--VAPGIG-----PGVAAAAAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 535
QY 391 NG-----THFNPLKKQGDVEVKAGEL-----LCEFD 415
Db 536 PGLGVAGVPGVGLGVAGVPGFAGAGDEGVRRSLSPELREGDPSSQHLPTSPSPRPGA 595
QY 416 IDAIIKAGYEVTTPIVVSNYKTKGTPVNTYGLGEIEAGANLLNVAKKEAVPA 466
Db 596 LAAAKAAYGAAPVGLGLGALGGVIGP-GVVGAGPAAAAAQAQAQAQAQAQAQAQAQAQA 645

RESULT 8
US-08-836-325-2
; Sequence 2, Application US/08836325
; Patent No. 6110672
; GENERAL INFORMATION:
; APPLICANT: Mandel, Gail
; APPLICANT: Haleboua, Simon
; APPLICANT: Borden, Laurence A.
; TITLE OF INVENTION: Peripheral Nervous System Specific
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Usin
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave., N. W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,325
; FILING DATE: 2-May-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: PCT/US95/14251
; FILING DATE: 02-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/482,401
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/334,029
; FILING DATE: 02-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0917.0240002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1011 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-836-325-2

```

Query Match 4.6%; Score 108; DB 3; Length 1011;
Best Local Similarity 20.7%; Pred. No. 0.055;
Matches 91; Conservative 67; Mismatches 168; Indels 114; Gaps 20;

QY	28	LFGLDVAQACYQGTVPVLVWVSWILATIEKFLHKRLKGTADFL-----ITPVLTLTLLG 81
Db	553	IFDLVNTQA-FDITIMVLICLNMTMMVK-----EGQTEMDYVYLHWINMVFIILFTG 605
QY	82	FLTFIAGPAMRWGDVLAHGLQGLDFGPGVGLFLGVYSPIVITGLHQSPFPILEL 141
Db	606	ECVLKLIS-----LRH-----YF--TVG-----WNILYFVVVILSVGMFLAEMIEK 646
QY	142	FNQGGSFIFATASMANIQAACLAFFFLAKSEKLLAGACSVSAVLGITPAIFGVNL 201
Db	647	Y-----FVSPTLFRVILARIGIRL-----IKGAKGIRLLFALMWSLPALFNIGL 694
QY	202	RLRWPFPI-----GIGTAA-----IGGALIALENIKAVALGAAGFLGVVS 241
Db	695	LLFLVMFIYAIFGMSNFAYVKKEAGINDMFNFTFGNSMICLFIQTTS--GMDGLLAPIL 753
QY	242	IDAP-----DMVMFLVCNAVTFFLAFGAAIAYGLVLRNGS 278
Db	754	NSAPDCDPKKVHPGSSVGDCNPSVGIFYFVSYIIISFLVVNMVIA---VILENS 809
QY	279	IDPDATAAPVPAGTTKAEAEAPAEFNSDSTIIQAPITGEAIALSSVSDAMFASGKLGCV 338
Db	810	VATEESTPELSEDDFMFYEVMEKEFPDAT-----QFIECKLSD--FAAALDPPLL 859
QY	339	AIVPTKGQLVS---PVSQKVVAPFSGHAFVARTKAEDGSNVDIL---MHIGEDTVNLN 391
Db	860	IAPKNVQLIAMDLPVSGDRHICLDILFAFYKRVLGE--GGMDSLRSQMEERFMSANPS 918
QY	392	GTHFNP-----LKKQGDVKA 407
Db	919	KVSEYEPITTTLLKRRQEEVSA 938

RESULT 9
US-08-836-325-12
; Sequence 12, Application US/08836325
; Patent No. 6110672

```

: GENERAL INFORMATION:
: APPLICANT: Mangel, Gail
: APPLICANT: Haleboua, Simon
: APPLICANT: Borden, Laurence A.
: TITLE OF INVENTION: Peripheral
: TITLE OF INVENTION: Sodium Cha
: TITLE OF INVENTION: X-ray Diff
: TITLE OF INVENTION: X-ray Desig

```

TITLE OF INVENTION: Therapeutic
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C
STREET: 1100 New York Ave., N. W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,325
FILING DATE: 2-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14251
FILING DATE: 02-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/482,401
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/334,029
FILING DATE: 02-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0917.0240002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1989 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-836-325-12

Query Match 4.6%; Score 108; DB 3; Length 1989;
Best Local Similarity 20.7%; Pred. No. 0.16;
Matches 91; Conservative 67; Mismatches 168; Indels 114; Gaps 20;

Qy	28	LFGLDVAQAGYQGTVPVLVWVSWILATIEKFLKBLKGTADFL-----ITPVLTLLLTG	81
Db	1508	IFDLVNTQA--FDITIVLICLVNMTVMVEK-----EQTEYMDVVLWINVFIILFTG	1560
Qy	82	FLTFIAGPMRWGVDLAHGLQGLYDRCGPVGGLLFGLVYSPVITVGLHQSFPPIELEL	141
Db	1561	ECVLKLTIS-----LRH-----YYF--TVG---WNLYFVVVILSVGMFLAEMIEK	1601
Qy	142	FNQGGSFIFATASMANIAGAACLAFFVLAKSEKLLAGASGVSAVLGITEPAIFGVNL	201
Db	1602	Y-----FVSPFLFVIRIARIGRLRL-----IKGAKGRTLFLFALMSLPALFNIGL	1649
Qy	202	RLRWPEFFI-----GIGTAA-----ICGALIALFNKAVKALGAAGFLGVWS	241
Db	1650	LLFLVMFIVAIFGMSFAYVKKEAGINDMFNETFGNSMICLFOITISA--GWDGLLAPIL	1708
Qy	242	IDAP-----DMVMFLCAVVTFFIAGAAATAGLYLVRNGS	278
Db	1709	NSAPPOCDPKVKHPGSSVEGDCGNPSVGIFVSVIIISFLVWNMYIA-----VILENFS	1764
Qy	279	IDPDATAAPVPAGTTKAEAEAPAESNDSTIIQAPLTCEATLSSVDAMPASCKLGSV	338
Db	1765	VAYEESTPELSEDDPMFVEYVEWKEKDPDAT-----QFTEFOKLSD--FAALADPPLL	1814
Qy	339	AIYVTRGQLVS-----PVSCKIVVAPSPGHAFVTRKAEODGNSVDIL---MHIGFDTVNLN	391

Db 1815 IAKPNKVLQIAMDLPVSGDRIHCLDLILFAFTRKVLGE--GGENDSLRQSEERFMSANPS 1873
QY 392 GTHFNP-----LKKQGEVKA 407
Db 1874 KVSVEPIITTTLKRKQEVSA 1893

RESULT 10
US-08-836-325-10
; Sequence 10, Application US/08836325
; Patent No. 6110672
; GENERAL INFORMATION:
; APPLICANT: Mandel, Gail
; APPLICANT: Haleboua, Simon
; APPLICANT: Borden, Laurence A.
; TITLE OF INVENTION: Peripheral Nervous System Specific
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave., N. W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,325
; FILING DATE: 2-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14251
; FILING DATE: 02-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/482,401
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/334,029
; FILING DATE: 02-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0917.024.0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-836-325-10

Query Match 4.5%; Score 107; DB 3; Length 1984;
Best Local Similarity 20.7%; Pred. No. 0.2;
Matches 91; Conservative 66; Mismatches 169; Indels 114; Gaps 20;

QY 28 LFGLDVAQAGYQGTVPVLVWVSHILATIEKFLKRLKGTADFL-----ITPVLTLTLTG 81
Db 1506 IFDLVTNQA-FDITIMVLCLNMTVMVEK-----EGQTEYMDYVLHWINMVFILFTG 1558
QY 82 FLTFIAIGPAMRWGVDLAHLGLOGLYDFGGPVGGLFLGLVYSPIVITGLHQSFPPIELEL 141

Db 1559 ECVLKLIS-----LRH-----YYF--TVGWNIEFVVVILSVGMFLA-----EM 1596
QY 142 FNQGGSEIFATASWANTAOGAACLAFFLAKSEKLGKLAGASGYSAVLGITTEPAIFGVNL 201
Db 1597 IEK-----YFVSPTLFRVIRLARIGRIURL-----IRKAGIRTLFLFALMMSLPALFNIGL 1647
QY 202 RLWPFPI---GIGTAA-----IGGALIALFNKAKAVALCAAGFLGVWS 241
Db 1648 LLFLVMEIYAIFGMSNFAYVYKKEAGINDMFENFETFGNSMCLFOITTSAGWDGULLAPIL 1706
QY 242 IDAP-----DMVNFVCAVVTFFIAFGAAIAYGLYLVRNRS 278
Db 1707 NSAPPDCDPKPKVHPGSSVEGDCGNPSVGFVFSYIIISFLVVNMYIA-----VILENFS 1762
QY 279 IDPDATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEALSSVSDAMFASCKLGSV 338
Db 1763 VATEESTPELSEDDFEMFYEWEKFDPAT-----QTFECKLSD--FAAALDPPLL 1812
QY 339 AIVPTKGQVLS---PVSGKIVVAFPSGHAFVTRKAEDGNSVDIL---MHIGFDTVNLN 391
Db 1813 IAKPNKVLQIAMDLPVSGDRIHCLDLILFAFTRKVLGE--GGENDSLRQSEERFMSANPS 1871
QY 392 GTHFNP-----LKKQGEVKA 407
Db 1872 KVSVEPIITTTLKRKQEVSA 1891

RESULT 11
5164180-6
; Patent No. 5164180
; APPLICANT: Payne, Jewel; Sick, August J.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE
; AGAINST LEPIDOPTERAN PESTS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/451,389
; FILING DATE: 14-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 353,860
; FILING DATE: 18-MAY-1989
; SEQ ID NO: 6;
; LENGTH: 1129
5164180-6

Query Match 4.5%; Score 106.5; DB 6; Length 1129;
Best Local Similarity 22.4%; Pred. No. 0.093;
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;

QY 249 MELVCAVVTFFIAFGAAIAYGLYLVRNRSID-----PDATAAPVPAG-----TTK 294
Db 380 LFVLDTGTEFSASLTADLPSTIY--RORGIVDSLDVIPPQDNSVPPARAGFSHRLSHVMTL 437
QY 295 AEA-----EAP-----AEFSN---DSTIIQAPLTGEALSSVSDAMFASCKLGS 336
Db 438 SQAAGAVYTLRAPTFSWRHSAEFSNLIPSSQITQIPLT-KSINLGSCTSVKGPGETGG 496
QY 337 GVAIVPTKGOLVSPVSKIVVAFPSGHAFVTRKAEDGNSVDILMHIGFDTVNLNTHFN 396
Db 497 DILRRTSPGQISTL---RVITAPLSQRVRYRY---ASTTNLQFHTSIDGRPNQGNFS 551
QY 397 PLKKQGEVKAAGELLCEFDIDAIAKAGYEVTTPIVVSNYKKTGPVNTYIGLGEIAGANLL 456
Db 552 ATWSSGNGLOSG-----SFTAGF--TTPF---NFSNGSSIFTLSAHVFNSG-NEV 596
QY 457 NVAKKEAVPA 466
Db 597 YIDRIEFVPA 606

RESULT 12
5188960-2
; Patent No. 5188960

; APPLICANT: PAYNE, JEWEL; SICK, AUGUST J.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATE ACTIVE
; AGAINST LEPIDOPTERAN PESTS, AND GENES ENCODING NOVEL
; LEPIDOPTERAN-ACTIVE TOXINS
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/451,261
; FILING DATE: 14-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 371,955
; FILING DATE: 27-JUN-1989
; SEQ ID NO:2:
; LENGTH: 1179
5188960-2

Query Match 4.5%; Score 106.5; DB 6; Length 1179;
Best Local Similarity 22.4%; Pred. No. 0.1;
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;
QY 249 MFLCAVVTETIAGAAATAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 294
Db 380 LFVLDGTEFFASLTADLPSTIY--RQGTVDLSLDVIPPQNSVPARAGFSHRLSHVTML 437
QY 295 AEA-----EAP-----AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 336
Db 438 SOAGAVTTLRAPTFWSHRHSAEFSNLIPSSQITQIPLT-KSINLGSCTSVKPGFTGG 496
QY 337 GVALVPTKQLVSPVSKIVVAFPSGHAFVTRKAEDGNSVDILMHIGFDVNLNGHTFN 396
Db 497 DILRTSPGQISTL---RVTTIAPLSQRYRIRY--ASTTNLQFHTSIDGRPINQGNFS 551
QY 397 PLKQGEVAGELLCEFDIDAIRAAGVEVTTPIVSNYKKTGPVNYGLGEIEAGANLL 456
Db 552 ATMSGNGNLSG-----SFRTAGF--TTFP---NFSNGSIFTLSAHFVNSG-NEV 596
QY 457 NVAKKEAVPA 466
Db 597 YIDRIEFVPA 606

RESULT 13
US-08-804-227C-3
; Sequence 3, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kunstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1864 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-3

Query Match 4.5%; Score 106.5; DB 2; Length 1864;
Best Local Similarity 21.7%; Pred. No. 0.2;
Matches 94; Conservative 45; Mismatches 159; Indels 135; Gaps 20;
QY 7 SLVNGYDVATMAAGEMPW-----SLGLDVAAGYOGYTVLPVLVSVNIIATIEKFLHK 61
Db 346 ALLATYGTART---AERPLWLGLSKSNIGHTQAAAGVAGVIMVLMAMR-----HG 392
QY 62 RL-----KGTADFLITPVLTLTLLTGLTFTTIAIGPAMRWGVDVLAHGLQG 105
Db 393 RLPRTLHVDRTTRVDWEKGGVRLTETV-----PWPGEAGEPRRAG 434
QY 106 LYDFGGPVGGLFLGLVSPVIVITGLHQSFPPFIELEFNQGSFIFATASMANIAQGAACL 165
Db 435 VSSEGA-----SGTNAHVVLVESPAGEPPAAGRPEDTGGAW---TVS---GRGPAAL 480
QY 166 AVFLAKSEKLGKLAGAGSVSAVLGITPEIPFVGNLRLRWPFPIGIGTAAIGGALI---A 222
Db 481 RAQAARLYDALTGTTGTGCGAGG-----AGPGTAEVAGALAHART 522
QY 223 LFNIKAVALGA-----AGFLGVVVSIDAPDMVLMVCAVW---TFFTAFG---AAIA 267
Db 523 AFRRAVVVLGNGRAELLAGLRELAEEHPGPRVVTGTATPATERRTAFLEFSQGSQORAGS 582
QY 268 YGLYLVRN-----GSDPD-----ATAAPVPAGTTKAEAPAEFSN 305
Db 583 RGLY--RRHPVFAALDEVCAALEPHLRLDLMFABPGSPAEPLDRTEFTQPALEAL 640
QY 306 DSTIIQAP---LTGEAIALSSVSDAMFASGKLGSVAIVTKGOLVSPVSGKIVVAFPS 361
Db 641 QTALFRLAEHHGLRAELCGHSGVEIAAAH---AAGVLTLDPDAARLVA-ARGRLMQALPA 696
QY 362 GHAF-AVTRKAEAD 373
Db 697 GGAAALRATAEE 709

RESULT 14
US-08-801-344-9
; Sequence 9, Application US/08801344
; Patent No. 6087140
; GENERAL INFORMATION:
; APPLICANT: Cameron, Douglas C.
; APPLICANT: Shaw, Anita J.
; APPLICANT: Altaras, Nedim E.
; TITLE OF INVENTION: MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL
; TITLE OF INVENTION: FROM SUGAR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dewitt Ross & Stevens S.C.
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,344
; FILING DATE:

```

; CLASSIFICATION:      435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REFERENCE/DOCKET NUMBER:    09820.037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:   608-831-2100
; TELEFAX:     608-831-2106
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH:      462 amino acids
; TYPE:         amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Pyridine nucleotide transhydrogenase, subunit B
; ORGANISMID: B
; US-08-801-344--9

Query Match           4.5%; Score 105.5; DB 3; Length 462;
Best Local Similarity 21.18; Pred. No. 0.029;
Matches 91; Conservative 66; Mismatches 123; Indels 151; Gaps 24;

QY 27 SLGLDVAQAQ-----QGTVLPVLV-----VSMLATIEKFL 59
Db 48 TIFGPD'GNVGWILLAMVIGGAIGIRLAKKVEMTEPVELVAILHSFGVLAALVGFNSYL 107
QY 60 HKRLKGADFLITPVLLTLTFTTAIGAIPMRWGVDVLAHQLGQLYDFGGPVGGLLF 119
Db 108 HH-D-AGMAPILVNHLTEVF-LG---IFIG-AVTFTGSVAF-----GKLCC 148
QY 120 LVYS-PIVITGLHQSPPIELELNQCGSFFATASMANIAQGRACLAAYFFLAKSEKLG 178
Db 149 KISKPLMLPNRHK-----MNTAALVSVFLLLVFVRTDSVGLQ- 187
QY 179 LAGASGVSAVLGIPEPA-IFGVNLRWRPFFIGYTAAIGGA----LIALEN-IKAVALG 232
Db 188 -----VLLALMTAIALVFWHL-----VASIGGADMPVVVSMLSXSGWAAA 230
QY 233 AAGFL-----GVVISIDAPDMVFLCAVAVTTFTAAGAAIAYGLYLVRNGSIDPDA 283
Db 231 AAGFMSLNDLLIVTGALVSSGAILSYIMCKAMN-RSFISVIAGGF-----GTDCSS 281
QY 284 TAAPVPAGT-TKAEARPAE-FSDNSTIIQAPLTGEATA-----LSSVSDAMFASG---K 333
Db 282 TGDDQEVEGEHRREITAETAELLKNSHSVIITPGYGMAVAQAQYPVAETEIKLRARGINVR 341
QY 334 LGSGVAIVPTKGQLVPSVSGKIVFAVFFGHAFARTKA-----EDGSNNDI 379
Db 342 FG-----IHPVAGR-----PGHMNVLLAEAkvPYDIVLEMDINDDFADTD 384
QY 380 LMHIGHF-DTVN 389
Db 385 VLVIGANDTVN 395

RESULT 15
US-09-498-599--9
; Sequence 9, Application US/09498599
; Patent No. 6303352
; GENERAL INFORMATION:
; APPLICANT: Cameron, Douglas C.
; APPLICANT: Shaw, Anita J.
; APPLICANT: Altaras, Nedim E.
; TITLE OF INVENTION: MICROBIAL PRODUCTION OF
; TITLE OF INVENTION: 1,2-PROPANEDIOL FROM SUGAR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dewitt Ross & Stevens S.C.
```

Search completed: March 21, 2002, 16:23:16
Job time: 147 sec.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: March 21, 2002, 16:19:04 ; Search time 42.58 Seconds

(Without alignments)
814.145 Million cell updates/sec

Title: US-09-604-231-2

Perfect score: 2363

Sequence: 1 MAMVPSLVNGYDVATMAA.....IEAGNLLNVAKKAVPATP 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

1:	/SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2363	100.0	468	22	AAB66707 C-glutaminc phosph
2	2363	100.0	661	22	AAG92650 C glutamicum prote
3	2332	98.7	661	22	AAB69080 Brevibacterium lac
4	1821	77.1	362	22	AAB66708 C-glutaminc phosph
5	468.5	19.8	683	22	AAG93207 C glutamicum prote
6	468.5	19.8	683	22	AAB66721 C-glutaminc phosph
7	298.5	12.6	648	22	AAG98282 Escherichia coli p
8	234	9.9	135	22	AAB66723 C-glutaminc phosph
9	233.5	9.8	167	20	AA120012 B. burgdorferi ant
10	232.5	9.8	583	21	AA153620 Amino acid sequenc
11	166.5	7.0	135	20	AA120013 B. burgdorferi ant

12	166.5	7.0	413	22	AAB66722 C-glutaminc phosph
13	161.5	6.8	126	18	AAW28051 Amino acid sequenc
14	156	6.6	449	21	AA181556 Streptococcus pneu
15	154	6.5	108	21	AAB40791 Human ORFX ORF555
16	151.5	6.4	658	21	AA181595 Human ORFX ORF555
17	137	5.8	572	21	AA181595 Streptococcus pneu
18	134	5.7	423	22	AA181595 S. epidermidis ope
19	133	5.6	48	18	AAW28158 Amino acid sequenc
20	128.5	5.4	806	20	AA121797 Alteromonas lipase
21	128.5	5.4	809	20	AA121796 Alteromonas lipase
22	128	5.4	629	22	AAG92331 C glutamicum prote
23	128	5.4	629	22	AAB76821 Corynebacterium gl
24	118.5	5.0	805	22	AAG90449 C glutamicum prote
25	118	5.0	461	20	AA138792 Neisseria meningit
26	117	5.0	461	20	AA138791 Neisseria meningit
27	116	4.9	397	22	AAW76823 Corynebacterium gl
28	116	4.9	502	20	AAW99600 B-stationis multi-
29	116	4.9	506	21	AA181620 Streptococcus pneu
30	115	4.9	521	21	AA181578 Streptomyces globi
31	115	4.9	521	21	AA1813605 Streptomyces globi
32	114	4.8	461	20	AA138793 Neisseria gonorrhoe
33	114	4.8	461	20	AA138794 N. gonorrhoeae ant
34	113.5	4.8	1176	14	AA138755 Delta endotoxin.
35	112	4.7	1951	22	AA138755 Human adult form o
36	112	4.7	1951	22	AA138755 Human neonatal for
37	111	4.7	551	19	AAW98519 H. pylori GPO 132
38	110.5	4.7	528	22	AAW38697 Human polypeptide
39	110.5	4.7	581	22	AAW40483 Human polypeptide
40	109	4.6	3033	14	AA138755 NAMH virus strain
41	108.5	4.6	730	19	AAW46315 Human elastin cont
42	108.5	4.6	731	21	AA169068 Amino acid sequenc
43	108.5	4.6	731	22	AA169068 Human elastin prot
44	108.5	4.6	733	15	AA158653 Synthetic human tr
45	108.5	4.6	733	20	AA158653 Amino acid sequenc

ALIGNMENTS

RESULT 1

AA169068

ID AAB66707 standard; protein; 468 AA.

XX AAB66707;

XX AC AAB66707;

XX DT 09-APR-2001 (first entry)

XX DE C-glutaminc phosphoenolpyruvate protein #1.

XX KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

XX OS Corynebacterium glutamicum.

XX PN WO200102583-A2.

XX PD 11-JAN-2001.

XX PF 27-JUN-2000; 2000WO-IB00973.

XX PR 01-JUL-1999; 99US-0142691.

XX PR 23-AUG-1999; 99US-0150310.

XX PR 03-SEP-1999; 99DE-1042095.

XX PR 03-SEP-1999; 99DE-1042097.

XX (BADI) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habehauer G;

XX WPI; 2001-080989/09.

XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate; sugar phosphotransferase system proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers

PT for transformation -

XX Claim 4; Page 101-102; 144pp; English.

XX The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.

XX Sequence 468 AA;

Query Match 100.0%; Score 2363; DB 22; Length 468;
Best Local Similarity 100.0%; Pred. No. 7.7e-228;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMVPSLVNGYDVAATMAAGEMPMSLFGDLVAQAGYQGTVPVLVSVWILATIEKFLH 60
DB 1 mamvpslvngydvaaatmaagempmslfgldvaagyggtvlpvlvsvwilatiekflh 60
QY 61 KRKLGTADEFLITPVLTLLLTGFTFTFAIGPAMRWGVDVLAHGLQGLYDFGPGVGGLLFGL 120
DB 61 krkgtadflitpvtllltgtftftaigpamrwvgdvdlahglqglydfgpgvggllfgl 120
QY 121 VYSPVITGLHQSFPPPELLEFNQGSFIFATASMANIAGAACLVFFLAKSEKLGKLA 180
DB 121 vvspvittglhqsfppllelefnqgsfifatasmnialagaaclavfflakseklkgl 180
QY 181 GASGVSALVIGTPEAFGLVNLRLRWPFFIGTAAIGGALIALFNKAVAGFLGVV 240
DB 181 gasgvsavigtpeaifgvnlrlrwpffigtaaiaggaliafnkavagflgvv 240
QY 241 SIDAPDMVFLCAVVTFFTAAGAAAYGLYLVRRNGSIDPDATAPVPAGTTKAEAP 300
DB 241 sidapdmvflcavvtfftaagaaayglylvrrngsidpdatapvpagttkaeap 300
QY 301 AEFNSNSTIIQAPLTGEATLSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFP 360
DB 301 aefnsnstiiqapltgeatalsvdsamfasklgsgvaivptkgqlvspvsgkivvafp 360
QY 361 SGHAFVTRKAEKGSNVDILMHIGFTVNLNGTHFNPLKKQGEVKAAGELLCFEDIDA 420
DB 361 sghafvtrkaedgsnvdilmhigftvnlngthfnplkkqgevkagellcfedidaik 420
QY 421 AAGYEVTTPVVSNNYKKTGPVNTYGLGEIDAGANLLNVAKEAVPATP 468
DB 421 aagyevttpivvsnykktgpvntyglgeidaganllnvakeavpatp 468

RESULT 2
AAG92650
ID AAG92650 standard; Protein; 661 AA.

XX AC AAG92650;

XX DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6404.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.

OS Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

XX

PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI; 2001-376931/40.

XX N-PSDB; AAH67869.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 661 AA;

Query Match 100.0%; Score 2363; DB 22; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.3e-227;
Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAMVPSLVNGYDVAATMAAGEMPMSLFGDLVAQAGYQGTVPVLVSVWILATIEKFLH 60
DB 194 mamvpslvngydvaaatmaagempmslfgldvaagyggtvlpvlvsvwilatiekflh 253
QY 61 KRKLGTADEFLITPVLTLLLTGFTFTFAIGPAMRWGVDVLAHGLQGLYDFGPGVGGLLFGL 120
DB 254 krkgtadflitpvtllltgtftftaigpamrwvgdvdlahglqglydfgpgvggllfgl 313
QY 121 VYSPVITGLHQSFPPPELLEFNQGSFIFATASMANIAGAACLVFFLAKSEKLGKLA 180
DB 314 vvspvittglhqsfppllelefnqgsfifatasmnialagaaclavfflakseklkgl 373
QY 181 GASGVSALVIGTPEAFGLVNLRLRWPFFIGTAAIGGALIALFNKAVAGFLGVV 240
DB 374 gasgvsavigtpeaifgvnlrlrwpffigtaaiaggaliafnkavagflgvv 433
QY 241 SIDAPDMVFLCAVVTFFTAAGAAAYGLYLVRRNGSIDPDATAPVPAGTTKAEAP 300
DB 434 sidapdmvflcavvtfftaagaaayglylvrrngsidpdatapvpagttkaeap 493
QY 301 AEFNSNSTIIQAPLTGEATLSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFP 360
DB 494 aefnsnstiiqapltgeatalsvdsamfasklgsgvaivptkgqlvspvsgkivvafp 553
QY 361 SGHAFVTRKAEKGSNVDILMHIGFTVNLNGTHFNPLKKQGEVKAAGELLCFEDIDA 420
DB 554 sghafvtrkaedgsnvdilmhigftvnlngthfnplkkqgevkagellcfedidaik 613
QY 421 AAGYEVTTPVVSNNYKKTGPVNTYGLGEIDAGANLLNVAKEAVPATP 468
DB 614 aagyevttpivvsnykktgpvntyglgeidaganllnvakeavpatp 661

```
RESULT 3
ARB69080
ID AAB69080 standard; Protein; 661 AA.
XX
AC AAB69080;
XX
DT 20-APR-2001 (first entry)
XX
DE Brevibacterium lactofermentum sucrose PTS enzyme II protein SEQ ID NO:2.
XX
KW Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;
KW phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;
KW coryneform bacterium; phosphoenolpyruvate-sugar transport system.
XX
OS Brevibacterium lactofermentum.
XX
PN WO200102584-A1.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-JP04348.
XX
PR 02-JUL-1999; 99JP-0189512.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;
XX
DR WPI; 2001-138150/14.
XX
DR N-PSDB; AAF32543.
XX
PT Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT II obtained by cassette ligation-mediated amplification of downstream
PT domain of coryneform bacterium sucrose gene, with sucrose-binding
PT activity .
XX
XX
XX Claim 1; Page 29-32; 45pp; Japanese.
XX
XX The present sequence represents the Brevibacterium lactofermentum
XX sucrose PTS (phosphoenolpyruvate:carboxylate phosphotransferase system
XX or phosphoenolpyruvate-sugar transport system) enzyme II, which has
XX sucrose-binding activity. A coryneform bacteria produced with the
XX sucrose PTS enzyme II gene can have more efficient sugar uptake, and
XX improved amino-acid and nucleic acid productivity. The sucrose PTS gene
XX and it's disrupted gene, such as one without the sucrose PTS function,
XX can be used to produce new breeds of coryneform bacterial strains to
XX uptake sugar more efficiently e.g. glucose only or and sucrose, and can
XX have improved amino-acid and nucleic acid productivity.
XX
XX Sequence 661 AA;
XX
Query Match 98.7%; Score 2332; DB 22; Length 661;
Best Local Similarity 98.5%; Pred. No. 1.6e-224;
Matches 461; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 MAWVPSLVNGYDVAATMAAGEMPMSLFCGLDVAQAGYQGTVPVLVWSWILATIERFLH 60
DB 194 mamvftllvngydvaaatmagempmslflgdvaqagygqgtvlpvlvswilatiekflh 253
QY 61 KRKGTADFLITPVLTLTGTFTTAIGPAMRWGDVLAHQGLYDGGPVGGLLFLGL 120
DB 254 krlmgtdadflitpvltlitgtfttaigpamrwgdvllahg lqglydfgpgvggllflgl 313
QY 121 VYSPVITGLHQSFPPIELFNQGGSFIFATASMANIAQGAACLAFFELAKSEKLGLA 180
DB 314 vyspivitglhgsfbbpielelnqgg9sfifatasmanlaqgaacclavflakseklkgl 373
QY 181 GASGSAVLGITEPAIFGVNLRWPFFFIGTAAIGGALIALFNKAKVALGAAGFLGVV 240
DB 374 gasgsaavlgitepaifgvnlrwpffiygigttaaiggalialfndikavalgaagflgvv 433
```

```
QY 241 SIDAPDMVMFLCAVVTFFIAFGAAIAYGLYLVRNGSIDPDATAAPVPAGTTKAEAEAP 300
DB 434 sidapdmvmflcavvtfvfifagaaiaaygylvrrngsidpataapvpagttkaeaeap 493
QY 301 AEFSDNSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAF 360
DB 494 aefsdnstiiqapltgeaialssvdamfasklgsgvaivptkgqlvsvsgkivvafp 553
QY 361 SGHAFVTRKAEQSNVDILMHIGFDTVNLNGTHFNPLKKGDEVKAGELLCEDIDAIAK 420
DB 554 sg hafavtrkaedgsnvdiilmhigfdt vnlngthfnplkkgdevkagellcefdidaik 613
QY 421 AAGYEVTTPIVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPATP 468
DB 614 aagyevttplvsnynyktgpvntyg lgeieaganllnvakkeavpatcp 661

RESULT 4
AAB66708
ID AAB66708 standard; protein; 362 AA.
XX
AC AAB66708;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate protein #2.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
PA (BADI ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-080989/09.
XX
PT Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation .
XX
XX Claim 6; Page 104-106; 144pp; English.
XX
XX The present invention relates to Corynebacterium glutamicum
XX phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.
XX The PTS nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria, the typing or identification of C. glutamicum or
XX related bacteria, as reference points for mapping C. glutamicum genome,
XX and as markers for transformation.
XX
XX Sequence 362 AA;
XX
Query Match 77.1%; Score 1821; DB 22; Length 362;
Best Local Similarity 100.0%; Pred. No. 9.6e-174;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 107 YDFGCPVGGLLFGLVSPVITGLHQSFPPIELFNQGGSFIFATASMANIAQGAACLA 166
DB 107 ydfgcpvggllfglvsvpivitglhqsfpfpielelnqgg9sfifatasmaniaqgaaccla 166
```

Db 1 ydfgppvgllflgvyspivltghqsfppielelfnfgggsffifatasmaniaggacla 60
Qy 167 VFFLAKEKLGAGASVAVLGTEPAIFGVNLRWRPFEGTGAAGGALIALFNI 226
Db 61 vfflaksekklglagagvsavlgitepaifgvnlrwrpffigtaaggaialfni 120
Qy 227 KAVALGAAGFLGVVSDIAPDMVFLVCAVTFVFFIAFGAAIAYGLVLRNGSIDPDATAA 286
Db 121 kavalgaagflgvsvdsdpdmvmfvcavtffiafgaaiayglylvrrngsidpataa 180
Qy 297 PVPAGTTKAEAEAPAEFSNDSTIIOTAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346
Db 181 pvpagttkaeaeapaeFSNDSTIIOTAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
Qy 347 LVSPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFDVNLNGTHFNPLKKGDEVK 406
Db 241 lvspvsgkivvafpsghafvtrkaedgsnvdilmhigfdvnlngthfnplkkgdevk 300
Qy 407 AGELLCFDDIDAIRAAGYEVVTPPIVWSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
Db 301 agellcefdidairaagyeVTPPIVWSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 360
Qy 467 TP 468
Db 361 tp 362
RESULT 5
AAG93207
ID AAG93207 standard; Protein; 683 AA.
XX AC AAG93207;
XX DT 26-SEP-2001 (first entry)
XX DE C glutamicum protein fragment SEQ ID NO: 6961.
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis.
XX OS Corynebacterium glutamicum.
XX PN EP1108790-A2.
XX PD 20-JUN-2001.
XX PF 18-DEC-2000; 2000EP-0127688.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX PA (KYOW) KYOWA HAKKO KOGYO KK.
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
XX N-PSDB; AAH68426.
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS Claim 29; SEQ ID NO: 6961; 246pp + Sequence Listing; English.
XX CC The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX SQ Sequence 683 AA;
Query Match 19.8%; Score 468.5; DB 22; Length 683;
Best Local Similarity 29.8%; Pred. No. 6.3e-38;
Matches 141; Conservative 79; Mismatches 206; Indels 47; Gaps 9;
Qy 27 SLFGLDVAQAGYQGTVPVLVSVILATIERFLKRLKGTADFLITPVLTLGLLGLFLFI 86
Db 218 tvfglpmvlnndysgvpfppllaaiglyvwekgklkilpeavgmfvfssllimipataf 277
Qy 87 AIGPAMRWGVGDVLAHGLQGLYDFGPGVGLLFGVYSPIVITGLHOSFPPIELELFNQ-G 145
Db 278 llgpfvgvngisnllleainmfpsfiisvlpilpvlplghlnpinaimiqnintlg 337
Qy 146 GSFIFATASMANIAQGAACLAFFFLAKSEKLGLAGAS--GVSAVL--GITEPAIFGVNL 201
Db 338 ydfigppmgawnfacfgivtgvfllsikernkamqvslgmlagllggisepslygvll 397
Qy 202 RLRPFFIGTAAIGGALIALFNKKAVALGAAGFLGVVSDIAPDMVFLVCAVTFPIA 261
Db 398 rfkkytyrlpgclagivgmifdikayafvtslltipand--pwlgytigiavafivs 455
Qy 262 FGAAIAYGLVLRNGSID-----PDAT-AAPV----- 288
Db 456 mflvliald---yrnsneerdearakaadkqeedlkaeanatpaapvaagagagaga 512
Qy 289 PAGTTKAEAEAPAEFSNDSTIIOTAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGOLV 348
Db 513 aagaataavaakpklaagevdivsplegkaiplsevpdpifaegklpggiaiqtgntvv 572
Qy 349 SPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFDVNLNGTHFNPLKKGDEVK 408
Db 573 apadatvllvqksghavairl-----dsgveillvhvgtlvqggefthvverrqvkag 628
Qy 409 ELLCEFDIDAIRAAGYEVVTPPIVWSNYKKTGPVNTYGLGEIEAGANLLNVAKK 461
Db 629 dplitfdadfirskdlpitpvvsvnaakfgeiegipadqanssttvikvngk 681
RESULT 6
AAB66721
ID AAB66721 standard; protein; 683 AA.
XX AC AAB66721;
XX DT 09-APR-2001 (first entry)
XX DE C.glutamicum phosphoenolpyruvate protein #5.
XX KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX OS Corynebacterium glutamicum.
XX PN WO200102583-A2.
XX PD 11-JAN-2001.
XX PF 27-JUN-2000; 2000WO-IB00973.
XX PR 01-JUL-1999; 99US-0142691.
XX PR 23-AUG-1999; 99US-0150310.
XX PR 03-SEP-1999; 99DE-1042095.
XX PR 03-SEP-1999; 99DE-1042097.

PA (BADI) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-080989/09.
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation -
XX Claim 32; Page 136; 144pp; English.
XX The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins which are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.
XX Sequence 683 AA;

Query Match 19.8%; Score 468.5; DB 22; Length 683;
Best Local Similarity 29.8%; Pred. No. 6.3e-38;
Matches 141; Conservative 79; Mismatches 206; Indels 47; Gaps 9;
QY 27 SLFGLDVAQAQYOGTVLPVLVSWILATIEKFLHKLKGTADFLITPVLTLTLTGFLTFI 86
Db 218 tvfqlpmvindyqgvpfpilaiaiglywvekglkkltpaeavqmvfpffallimipataf 277
QY 87 AIGPAMRWGVDVLAHGLQGLYDGGVPVGLLFLGVSPVITGLHOSFPPIELFNQ-G 145
Db 278 llqpfvgigvngisnlleaonfspfillsivlpflvplglhwpnlaimiqnintlg 337
QY 146 GSIFATASMANIAOQAACLAFLAKSEKLGLAGAS--GVSANVL--GITEPAIFCVNL 201
Db 338 ydfiqpmpgawnfacfglvtgvtglvflslkernkamrqsllgmilgglgisepslygvll 397
QY 202 RLWRFPGFIFIGTAAIGGALITALEFNKAVALGAAGFLGVWSIDAPDMVMFLVCAVWVFFIA 261
Db 398 rfkktfrrlpgclagivmgfidkayafvftsltipamd--pwlgtyiglavaffvs 455
QY 262 FGAAIAYGLVLRNRSID-----PDAT-AAPV----- 288
Db 456 mflviald---yrsneerdearakaadkqaedlkaeanatpaapvaagagagagaga 512
QY 289 PACTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKQOLV 348
Db 513 aegaatavaakpklageevdvlsplegkaipisevpdpifaagkigpgiaigtntv 572
QY 349 SPVSGKIVVAFPSGHAFVARTKAEDGNSVDILMHIGFDTVNLNGTHFNPLKKGDEVKAG 408
Db 573 apadatvilvqkshgavarl-----dsgevllvhvldvtqlvggegtvhrvrrqvkag 628
QY 409 ELLCEFDIDAKAAGVEYTPPIVVSNNYKKTGPVNTYGLGIEIAGANLLNVAKK 461
Db 629 dplittfdaadfirskdpllipvsvsnaakfgelegipadqanssttikvngk 681

RESULT 7
AAG98282
ID AAG98282 standard; Protein; 648 AA.
XX
AC AAG98282;
XX
DT 21-SEP-2001 (first entry)
XX
DE Escherichia coli protein sequence SEQ ID NO:330.
XX
KW Escherichia coli; identification; proliferation; microorganism;

KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition.
XX Escherichia coli.
OS WO200148209-A2;
XX
PN 05-JUL-2001.
PD
XX 19-DEC-2000; 2000WO-US34419.
PF
XX 23-DEC-1999; 99US-0173005.
PR
XX (ELIT-) ELITRA PHARM INC.
PA
XX Forsyth RA, Ohlsen KL, Zyskind JW;
XX WPI; 2001-457376/49.
DR N-PSDB; AAH81338.
DR
XX Novel nucleic acids encoding proteins required for Escherichia coli
PT proliferation, useful for screening for antimicrobial agents -
PT
XX Claim 19; Page 462-463; 596pp; English.
XX
CC The present invention describes a purified or isolated nucleic acid
CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (I) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.
CC For example, nucleic acid probes complementary to proliferation-required
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.
XX Sequence 648 AA;

Query Match 12.6%; Score 298.5; DB 22; Length 648;
Best Local Similarity 21.7%; Pred. No. 6.1e-21;
Matches 134; Conservative 82; Mismatches 214; Indels 187; Gaps 17;
QY 13 DVAAATMAAGEMPWMLFGLDVAQAQYOGTVLPVLVSWILATIEKFLHKLKGTADFLIT 72
Db 51 nlalifaigvassv---kdsagaalagavgyfvltkamvtinpein---mgvlaglit 104
QY 73 PVL-----TILLTGFLTF-----IAIGPAMRWGVDVLAHGLQGLYD 108
Db 105 glvggaaynrwsdklpldfisffggkrfvpiaatgfcflvaiaaifgyvppvqhahagge 164
QY 109 F---GGPVGGLLGLVYSPVITGLHOSFPPI-----EELFNQO-- 145
Db 165 wlvsgalgsigfignrlilptglhqlvntiawfigeftnaagtvfhgdinrfyagdg 224
QY 146 --GSFIFATASMANIAOQAACLAFLAVFFLAKSEK---LKLGLAGASCVSAVL-GITEPAIF-- 197
Db 225 tagmfmsgffimmfmlpgaalamyfaapkrpvmvgmllsvavtafltgvtpleflf 284
QY 198 -----GVNRLRWPFFIFIGITAAIGGAL----- 220
Db 285 mflapllyllhalltgislftvatllghagfsagaidyalmynlpaasqnvmlvmg 344

Qy 221 -----IALFNKA-----VALGAA- 234
| :||:|
Db 345 viffayfvfslvirmfoltgptgredkedelivteeansteegltqlatnyiaavgtd 404
| :||:|
Qy 235 -----GFLGVWSIDAPDMWFLVCAVVTFFIAFGA 264
| :||:|
Db 405 nlkaidacitrlrltvadsarvndtmckrlgagvgvklmkgtlqivgakaesigdamkk 464
| :||:|
Qy 265 AIAYGLYLVRNGSIDPDATAPVPAGTTKAEAPAEFSDSTIIQAPLTGEATLSSV 324
| :||:|
Db 465 vvar9-pvaasaeatp-ataavakpavpnasiae-----ivspitgvdvvaldqv 515
| :||:|
Qy 325 SDAMFASGLSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFARVTKAEDGNSVDILMHIG 384
| :||:|
Db 516 pdeafaskavgdvavkptdkivvpaagtivkifntnhafclele---kgaievvhmg 571
| :||:|
Qy 385 FDTVNLNGTHFNPLKKQDEVKAGELLCEFDIDAIRAAGYEVVTPPIVVSNNYKKTGPVNTY 444
| :||:|
Db 572 idtvaldegkfriveegavsqagqpillemdlylnanarismispvvcnsiddfsgliik 631
| :||:|
Qy 445 GLGEIEAGANLLNVAKK 461
| :||:|
Db 632 aghivagqtplyeikk 648
| :||:|

RESULT 8

AAB66723
ID AAB66723 standard; protein; 135 AA.

AC AAB66723;

XX 09-APR-2001 (first entry)

XX C.glutamicum phosphoenolpyruvate protein #7.

XX Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

XX Corynebacterium glutamicum.

XX WO200102583-A2.

XX 11-JAN-2001.

XX 27-JUN-2000; 2000WO-IB00973.

XX 01-JUL-1999; 99US-0142691.

XX 23-AUG-1999; 99US-0150310.

XX 03-SEP-1999; 99DE-1042095.

XX 03-SEP-1999; 99DE-1042097.

XX (BADI) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-080989/09.

XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;
XX sugar phosphotransferase system proteins or their portions, useful for
XX typing or identifying C. glutamicum or related bacteria, and as markers
XX for transformation -

XX Claim 36; Page 142-143; 144pp; English.

XX The present invention relates to Corynebacterium glutamicum

XX phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.

XX The PTS nucleic acids and proteins are useful in the

XX identification of microorganisms which can be used to produce fine

XX chemicals, for modulating fine chemical production in C. glutamicum or

XX related bacteria, the typing or identification of C. glutamicum or

XX and as markers for transformation.

XX Sequence 135 AA;

Query Match 9.9%; Score 234; DB 22; Length 135;

Best Local Similarity 37.5%; Pred. No. 1.7e-15;

Matches 51; Conservative 25; Mismatches 56; Indels 4; Gaps 1;

Qy 326 DAMFASGLSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFARVTKAEDGNSVDILMHIGF 385
| :||:|
Db 2 dpifaagklpggiaiqtptgntvpadatvilvqkshavairl----dsgveilvhvql 57
| :||:|

Qy 386 DTVNLNGTHFNPLKKQDEVKAGELLCEFDIDAIRAAGYEVVTPPIVVSNNYKKTGPVNTY 445
| :||:|

Db 58 dtvqlggsegftvhvrrgqvkagdpilftdadirskdlpitpvvsvnaakfgelegip 117
| :||:|

Qy 446 LGEIEAGANLLNVAKK 461
| :||:|

Db 118 adqanssttvikvngk 133
| :||:|

RESULT 9

AAY20012
ID AAY20012 standard; protein; 167 AA.

XX AAY20012;

XX 19-JUL-1999 (first entry)

XX B. burgdorferi antigenic protein, f346.aa.

XX Antigenic protein; vaccine; Lyme disease; infection; detection.

XX Borrelia burgdorferi.

XX WO9859071-A1.

XX 30-DEC-1998.

XX 18-JUN-1998; 98WO-US12718.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (MEDI-) MEDIMUNE INC.

XX Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI; 1999-189980/16.

XX N-PSDB; AAX61709.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the diagnosis, prevention and treatment of diseases
XX caused by Borrelia, particularly Lyme disease

XX Claim 12; Page 161; 275pp; English.

XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX can be used in vaccines for eliciting protective antibodies to members of
XX the Borrelia genus, particularly for the use against Lyme disease in
XX humans and animals. They can be used for preventing or attenuating an
XX infection caused by a member of the Borrelia genus. The products can also
XX be used for detection of members of the Borrelia genus.

XX Sequence 167 AA;

Query Match 9.9%; Score 233.5; DB 20; Length 167;

Best Local Similarity 34.8%; Pred. No. 2.6e-15;

Matches 46; Conservative 30; Mismatches 51; Indels 5; Gaps 2;

PT	products for the diagnosis, prevention and treatment of diseases
PT	caused by <i>Borrelia</i> , particularly Lyme disease
xx	
xx	Claim 12; Page 161; 275pp; English.
xx	
CC	This sequence represents a <i>Borrelia burgdorferi</i> (Bb) protein of the
CC	invention, which is suitable for use in a vaccine. The Bb polypeptides
CC	can be used in vaccines for eliciting protective antibodies to members of
CC	the <i>Borrelia</i> genus, particularly for the use against Lyme disease in
CC	humans and animals. They can be used for preventing or attenuating an
CC	infection caused by a member of the <i>Borrelia</i> genus. The products can also
CC	be used for detection of members of the <i>Borrelia</i> genus.
xx	
xx	Sequence 135 AA;
SQ	

```
Query Match      7.0% Score 166.5; DB 20; Length 135;  
Best Local Similarity 34.7%; Pred. No. 9.7e-09;  
Matches 34; Conservative 21; Mismatches 38; Indels 5; Gaps  
  
QY   353 GKIVAPFSGHAFVAVRYKAEDGSNDILMHGFTVTNLNGTHFNPLKKQGDEVKAGELLCC 412  
||||| : ||||| : ||||| : ||||| : ||||| :  
Db    3 gkgikfktnhafaletk---egveifvfhgintlnngkftvrvaeeingvkagevii 58  
  
QY   413 EFDIDAKAACYEYTTPIVVSNNYKKTGPVNYYGLGEIE 450  
|:::| |:::| ::||:| :|::| :|:  
Db    59 rldleylkehsevitpwiansdevsie-vsfarle 95
```

RESULT 12
AAB66722
ID AAB66722 standard; protein; 413 AA.

DT 09-APR-2001 (first entry)
XX AAB66722;

DE C.glutamicum phosphoenolpyruvate protein #6.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum

PN WO200102583-A2.

11-JAN-2001.

27~JUN-2000: 2000WO-TB00973.

PR 01-JUL-1999: 99US-0142691.

EK Z3-AUG-1999; 99US-0130310.
 PR 03-SEP-1999; 99DE-1042095

PR 03-SEP-1999; 99DE-I042097.
XX

PH (BADT) BASE AG.
XX

XX
 DR WPI: 2001-080989/09.
 PFI Pompejus M, Kroeger B, Schroeder H, zelder O, Haberhauer G;

Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate: sugar phosphotransferase system proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers for transformation -

PS Claim 34; Page 139-140; 144pp; English.

The present invention relates to *Corynebacterium glutamicum* phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins. The PTS nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or

CC related bacteria, as reference points for mapping *C. glutamicum* genome, CC and as markers for transformation.

Sequence 413 AA:

Query Match	7.0%	Score 166.5;	DB 22;	Length 413;
Best Local Similarity	28.9%	Pred. No. 5.1e-08;		
Matches	54;	Conservative	35;	Mismatches 93;
			Indels	55;
			Gaps	3;

Qy	27	SIFGLDVAQAGYQGFWLPVLVVSWILATIEFLHKRLKGTADFLLTPVLTLLLTGFLTFI	86
Db	218	tvfglpmvIndysgavfppliaaiglywvekgllkiipeavqmvfvfslIimipataf	277

Qy	87	AIGPAMRWGDVLAHGLQGGLYDFGGPVGGLFGLVSPITVITGLHQSFPPTELELFNQ-G	145
Db	278	Llqpfcfagvngaisllleaiaofsfilsivallvoflvglahwalmianintla	337

Qy 146 GSFIFATASMANIAQGAACIAVFFLAKSEKUKLAGAS--GVSAVL--GITEPAIFGVNL 201

Db 338 vdfiagomcawnfafca]vtravfl]s[kernkamravs]aam]aql]ant]seng]vov]] 307

Ov 202 RI.RWPFF 208

db 398 r f k k t v f 404

RESULT 13

AAW28051
ID AAW28051 standard: Protein: 126 AA.

AA AAW28051:

27-AUG-19

XX Amino acid sequence of sta

XX Staphylococcus aureus protein: ribozyme: antisense se
KW

KW staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome.

OS Staphylococcus aureus.

AA	FH	Key	Location/Qualifiers
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

```

FT      misc difference 1.120
FT      /note= "residues designated X are not defined in
FT      the specification"

```

PN W09730070-A1.

21-AUG-1997.

AA
PF 19-FEB-1997: 97WO-US02318

XX
PR 20-FEB-1996: 96US-0011888XX
PA (SMTK) SMITHKLINE BEECHAM CORP

Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Pratt JM, Reichenberg M, Ward JM;
 PI Burnham MK, Reichenberg M, Ward JM;
 AA

WPI: 1997-424969/39

DK N-PSDB; AA184002.
XX

PT aureus infection
PT to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection
PT novel polypeptides) from *Staphylococcus aureus* strain WC0H29 -
PT used as adjuvants in vaccines against S.

Claim 6: Pages 432-433: 989pp: English

The present sequence represents a *Staphylococcus aureus* protein, that, based on homology with a human or *Staphylococcus aureus* protein, is

CC believed to be a nerve growth factor or p75 protein. The DNA sequence
 CC was isolated from a library of clones of *S. aureus* WCUH 29 in
 CC *Escherichia coli*. The DNA sequence can be used in the construction of
 CC ribozymes and antisense sequences to control the expression of
 CC staphylococcal genes. The DNA sequence is also useful as a source of
 CC regulatory elements for the control of bacterial gene expression. The
 CC present protein may be used to produce vaccines to enable a host to
 CC produce specific antibodies with antibacterial action. These vaccines
 CC and antibodies would protect a host against invasion by *S. aureus*, and
 CC conditions relating to staphylococcal infection, e.g. staphylococcal
 CC food poisoning, scaled skin syndrome, and toxic shock syndrome.

XX Sequence 126 AA;

Query Match 6.8%; Score 161.5; DB 18; Length 126;
 Best Local Similarity 32.8%; Pred. No. 2.8e-08;
 Matches 40; Conservative 25; Mismatches 44; Indels 13; Gaps 4;

QY 334 LGSGVAI-VPTKGQLVSPVSGKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFDVNLNG 392

Db 2 vgegataikaheeskiapnglismivpckhavg1--gsedg--vdvihgvkxvdieg 57

QY 393 THFNPLKKQGDVVKAGELICEFDIDAIRAAGYEVTTPIVWSNYKKTGPVNTYGLGEIEAG 452

Db 58 kqkyfvnqndhveagqxlqfdqyiqqygynadxivvis-----nsadlkgkvelt 109

QY 453 AN 454

Db 110 mn 111

RESULT 14

AA81556
 ID AA81556 standard; Protein; 449 AA.

XX AA81556;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae type 4 protein sequence #56.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.

XX Streptococcus pneumoniae.

OS WO200006737-A2.

PN 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

PR 19-MAR-1999; 99US-0125164.

XX (MTCR-) MICROBIAL TECHNIQS LTD.

PI Gilbert CFG, Hansbro PM;

XX WPI; 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein

PS Claim 1; Page 78; 108pp; English.

XX AA81501 to AA81679 represent specifically claimed protein sequences
 CC isolated from *Streptococcus pneumoniae*. AA805407 to AA805590 represent
 CC specifically claimed nucleotide sequences isolated from *S. pneumoniae*.
 CC The sequences have antibacterial and antiinflammatory properties.

CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of *S. pneumoniae* infection
 CC and meningitis. AA805591 to AA805614 represent primers used in the
 CC exemplification of the present invention.

XX Sequence 449 AA;

Query Match 6.6%; Score 156; DB 21; Length 449;

Best Local Similarity 25.0%; Pred. No. 6.5e-07;

Matches 85; Conservative 48; Mismatches 117; Indels 90; Gaps 17;

QY 1 MAWVPSLVNGVDVAATMAAGEMPMSLFLDVAQAQYQG-----TVLPVLV----- 48

Db 140 mmktvtslkeg---aviadg---wam-gnvvarfgtgitfaiimaivtviymcvk 190

QY 49 -SWILATIEKFLHKLKGTADFLITPVLLLTGTFLT-----FTAIGPAM----- 92

Db 191 hnwvikmpeavpegvsrg-----ftalvpgfvvatvifingllvamgtdfkvia 241

QY 93 -----RWGDVLAHGL-OGLYDFGGPVGGLFLGLVYSPIVITGLHQSFPIELE 140

Db 242 ipfgfvsnltswiglmilylltqlwivghganivfafv-spialanmaen----- 293

QY 141 LFNQGGSFIFA--TASMANIAOGAA-----CLAVFFLAKSEKLGAGASVAVLGITE 193

Db 294 --aagghfavagefsmfviaggsgatlgcliyafaskseqdkaigrasvvpalfine 351

QY 194 PAIFGV-----NLRLRWPFFIG-IGTAAIGGALIALNIKAV-----ALGAAFLGVV 240

Db 352 plifgpliylnpalaipfilapmvtatliyvanlfnfkpilaqvpwptvgigalgta 411

QY 241 SIDAPDMVMFLYCAVVTFFIAFGAAIAYGLYLVRNGSID 280

Db 412 dira--vlvalvcafaafliyfpfirvydqkvkveeqgiz 449

RESULT 15

AA840791

ID AAB40791 standard; Protein; 108 AA.

XX AAB40791;

XX 08-FEB-2001 (first entry)

DE Human ORF555 polypeptide sequence SEQ ID NO:1110.

XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

PD

XX

	PF	31-MAR-2000; 2000WO-US08621.	
	XX		
	PR	31-MAR-1999; 99US-0127607.	
	PR	02-APR-1999; 99US-0127636.	
	PR	05-APR-1999; 99US-0127728.	
	PR	30-MAR-2000; 2000US-0540763.	
	XX		
	PA	(CURA-) CURAGEN CORP.	
	XX		
	PI	Shimkets RA, Leach M;	
	DR	WPI; 2000-602362/57.	
	XX	N-PSDB; AAC75000.	
	PT	Novel nucleic acids and peptides derived from open reading frame X,	
	PT	useful for treating e.g. cancers, proliferative disorders,	
	PT	neurodegenerative disorders and cardiovascular disease -	
	PS	Claim 11; Page 1038; 5507pp; English.	
	XX		
	CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,	
	CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX	
	CC	sequences have activities such as: cytostatic; hepatotropic; vulnerary;	
	CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	
	CC	osteopathic; anticoagulant; antiarthritic; immunosuppressant;	
	CC	immunostimulant; cardiant; thrombotytic; coagulant; vasotrophic;	
	CC	anti-diabetic; hypotensive; dermatological; immunosuppressive;	
	CC	anti-inflammatory; antibacterial; antiviral; antifungal; antirheumatic;	
	CC	antithyroid; and antianaemic. The sequences can be used for determining	
	CC	the presence of or predisposition to, or preventing or treating	
	CC	pathological conditions associated with an ORFX-associated disorder. The	
	CC	nucleic acids can be used to express ORFX proteins in gene therapy	
	CC	vectors. The proteins and nucleic acids may be used to treat cancers,	
	CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,	
	CC	graft vs host disease, cardiovascular disease, diabetes mellitus,	
	CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus	
	CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,	
	CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,	
	CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,	
	CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance	
	CC	coagulation; to inhibit thrombosis; and as a contraceptive.	
	XX		
	SQ	Sequence 108 AA;	
		Query Match 6.5%; Score 154; DB 21; Length 108;	
		Best Local Similarity 32.7%; Pred. No. 1.2e-07;	
		Matches 32; Conservative 19; Mismatches 43; Indels 4; Gaps 1;	
QY	300	PABFSNDSTIIQAPLCEATLSSVSDAMPASGLAGSVAIVPTKGVLSPVSGKVVAF 359	
		: : : : : :	
Dd	14	pssalkpsllkivaplggilvpldqvpdpvfqakmvvgdglsldpisnellapvtgtqlh 73	
		: : : : : :	
QY	360	PSGHAFAVRTKAEDEGNVDILMHIGFETVMNLTHTNP 397	
		: : : : : :	
Dd	74	nahaitittp-----egievlvhigdtvmrlrgdsyp 107	

Search completed: March 21, 2002, 16:22:45
Job time: 221 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:23:20 ; Search time 17.08 seconds

(without alignments)
1004.633 Million cell updates/sec

Title: US-09-604-231-2

Perfect score: 2363

Sequence: 1 MANVFPISLVNGYDVAAATMAA.....IEAGANLLNVAKKEAVPATP 468

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1058.5	44.8	651	1	PTSA_PEDPE
2	953.5	40.4	664	1	PTSA_STRMU
3	701	29.7	480	1	PTSB_STAXY
4	678	28.7	474	1	PTSB_PASMU
5	657	27.8	609	1	PTSB_BACSU
6	646	27.3	479	1	PTSB_VIBAL
7	640	27.1	631	1	PTBA_ERWCH
8	570	24.1	470	1	PTTB_BACSU
9	567.5	24.0	625	1	PTBA_ECOLI
10	492	20.8	456	1	PTSB_SALTY
11	489	20.7	456	1	PTSB_KLEPN
12	462.5	19.6	674	1	PTBA_CORGL
13	406	17.2	459	1	SACK_BACSU
14	390	16.5	460	1	PTSB_BACSU
15	367	15.5	473	1	PTTB_ECOLI
16	321.5	13.6	699	1	PTGA_BACSU
17	317.5	13.4	651	1	PTAA_KLEPN
18	298.5	12.6	648	1	PTAA_ECOLI
19	294.5	12.5	324	1	PTGB_BACST
20	287.5	12.2	631	1	YBFS_BACSU
21	270.5	11.4	485	1	PTDA_ECOLI
22	256	10.8	189	1	PTGA_BORBU
23	255	10.8	168	1	YPOE_BACSU
24	254.5	10.8	634	1	LACY_STRTR
25	250.5	10.6	474	1	YFEB_ECOLI
26	249.5	10.6	154	1	PTGA_MYCCA
27	246	10.4	168	1	PTGA_ECOLI
28	244.5	10.3	482	1	PTSB_VIBCH
29	244	10.3	168	1	PTGA_SALTY
30	238.5	10.1	165	1	PTGA_HAEIN
31	233	9.9	627	1	LACY_LACDE
32	231.5	9.8	161	1	PTGA_BUCAL
33	230.5	9.8	940	1	PTGA_MYCPN

34	221	9.4	641	1	RAFP_PEDPE	P43466 pediococcus
35	204	8.6	908	1	PTGA_MYCGE	P47315 mycoplasma
36	165.5	7.0	639	1	LACY_LEULA	O48624 leuconostoc
37	159	6.7	483	1	PTVB_ECOLI	P32154 escherichia
38	151.5	6.4	658	1	HRSA_ECOLI	P54745 escherichia
39	144.5	6.1	527	1	PTIB_BACSU	P54715 bacillus su
40	136.5	5.8	359	1	PTWC_ECOLI	P32672 escherichia
41	130	5.5	578	1	PTFB_RHOCA	P23387 rhodobacter
42	126.5	5.4	580	1	PTFB_XANCP	P23355 xanthomonas
43	123.5	5.2	632	1	PTMA_BUCAL	P57635 buchnera ap
44	122	5.2	694	1	PTFA_MYCPN	P75039 mycoplasma
45	120.5	5.1	577	1	PTLB_LACCA	P24400 lactobacill

ALIGNMENTS

RESULT	1
PTSA_PEDPE	
ID	PTSA_PEDPE
AC	P43470; STANDARD; PRT; 651 AA.
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (ETIABC-SCR) (SUCROSE-
DE	PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE	COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN	SCRA.
OS	Pediococcus pentosaceus.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC	Pediococcus.
OX	NCBI_TaxID=1255;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=PEEL.0;
RA	Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL	Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
CC	-!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC	SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC	-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC	AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC	PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC	PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC	THE SUGAR.
CC	-!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =
CC	PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC	-!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC	-!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC	-!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; Z32771; CAA83668.1; -
DR	EMBL; L32093; AAA25567.1; -
DR	HSSP; P20166; IAX3.
DR	InterPro; IPR001127; PTS_EIIA.
DR	InterPro; IPR001996; PTS_EIIB.
DR	InterPro; IPR003352; PTS_EIIC.
DR	Pfam; PF00358; PTS_EIIA.1; 1.
DR	Pfam; PF00367; PTS_EIIB.1.
DR	Pfam; PF02378; PTS_EIIC.1.
DR	ProDom; PD001476; PTS_EIIB.1.
DR	ProDom; PD002243; PTS_EIIA.1.
DR	PROSITE; PS00371; PTS_EIIA.1; 1.
DR	PROSITE; PS01035; PTS_EIIB.CYS; 1.
KW	Phosphotransferase system; Sugar transport; Transferase;

```

KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40 EIIB DOMAIN.
FT DOMAIN ? ? EIIC DOMAIN.
FT DOMAIN 510 651 EIIA DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 562 562 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 651 AA; 68454 MW; C87BA09D550A77F8 CRC64;

Query Match 44.8%; Score 1058.5; DB 1; Length 651;
Best Local Similarity 46.7%; Pred. No. 1.2e-62;
Matches 221; Conservative 78; Mismatches 151; Indels 23; Gaps 5;

QY 1 MAMVPSLVNGYDYAATMAAGEMPMSLFGLDVAQAGVGQVPLVPLVSVWILATIEKFLH 60
DB 194 MIMVPLSVNGYSVATMAAGKQVYVNVFGLHVAQAGTQGVPLVGVAFILATLEKFFH 253
QY 61 KRLKGTADFLITPVLTLLLTGFLTFIAGFAMRWVGDVLAHGLQGLYDFGPGVGLLFLG 120
DB 254 KHKGAEDEFTPMFAIVITGFLFTIVGPVLRVSDALTNGLVLYNSTGWMGIFGL 313
QY 121 VSPVITVGLHQSPPPELELF-----NOGGSFIFATSMANIAQAACLAFLVFLAKSEKL 176
DB 314 LYSIVITGLHQTTPAETQTLLANVAKTGGSFIPVASMANGQGAATLAIFFAATKSQKQ 373
QY 177 KGLAGAGSVASVIGTIPAFVGNLRLRWPFIFIGTAAGIALGALIALFNKAVAGAGF 236
DB 374 KALTSSAGVSALLGITPAIFGVNLRKWFVFAAIAAGIASAFGLFHLVLSVAMGPSV 433
QY 237 LGVVSIDAPDMVFLVCAVVTFFTAAGAAIAYGLYLVRRNGSIDPDATAAPVPAGTTKAE 296
DB 434 IGFISIASKSTPAEMLSAVSVFVAFPTFIY----AKRTLGDRDQVKSPAPTSTV--- 486
QY 297 AEAPAEFNSNTIIQALPTGEALSVSDAMFASGKLGSGVAIVPTKGOLVSPVSGKIV 356
DB 487 -----INVNDEIISAPVTGASELSKQVNDQVFSAEIMGKGAALVPSDDQVAPADGVIT 540
QY 357 VAPFSGHAFVTRKAEGNSVDILMHIGFDVNLNGTHFNPLKQGDDEVKAGELLCEFDI 416
DB 541 VTYSHAYAGIKTTA----GAEILIHGLDVLNGLNGEHTFTNVQKGDVTHQGDLLGTFDI 596
QY 417 DAIAAGYEVPTPIVWSNKKGTGVNTYGLGEIEAGANL--LNVAKKEAVPAT 467
DB 597 AALKANYPDPVMLIVTNTANYANVERLKVNTVQAGEQLVALTAPASSVAAT 649

RESULT 2
ID PTSA_STRMU STANDARD; PRT; 664 AA.
AC P12655;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89123027; PubMed=2536656;
RA Sato Y., Poy F., Jacobson G.R., Kuramitsu H.K.;
RT "Characterization and sequence analysis of the scrA gene encoding
RT enzyme IIScr of the Streptococcus mutans
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
RL J. Bacteriol. 171:263-271(1989).
RN [2]
RP SEQUENCE OF 639-664 FROM N.A.

```

```

RC STRAIN=GS-5;
RX MEDLINE=93329360; PubMed=8336109;
RA Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RT "Isolation, characterization and sequence analysis of the scrK gene
RT encoding fructokinase of Streptococcus mutans.";
RL J. Gen. Microbiol. 139:921-927(1993).
CC -I- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -I- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -I- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22711; AAA26971.1; -.
DR EMBL; D13175; BAA02466.1; -.
DR PIR; B32243; B32243.
DR HSP; P08837; 2F3G.
DR InterPro; IPR001127; PTS_EIIB.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIB_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIB; 1.
DR PROSITE; PS00371; PTS_EIIB_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
DR Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40 EIIB DOMAIN.
FT DOMAIN ? 664 EIIC DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 331 331 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 585 585 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 664 AA; 69988 MW; 809E63E32281A9A1 CRC64;

Query Match 40.4%; Score 953.5; DB 1; Length 664;
Best Local Similarity 42.6%; Pred. No. 1e-55;
Matches 205; Conservative 89; Mismatches 156; Indels 31; Gaps 9;

QY 1 MAMVPSLVNGYDYAATMAAGEMP-----MMSLFGLDVAQAGTQVPLVPLVSVWILA 53
DB 194 MIMVAPGAANIIGLAANAPISKAATIGAYTGFWNIFGLHVTQASYTYQVIVPLVAVWLLS 253
QY 54 TIEKFLHRLKGTADFLITPVLTLLLTGFLTFIAGFAMRWVGDVLAHGLQGLYDFGPGV 113
DB 254 ILEKFLHRLPSAVDFTTPLLVIITGFLTFIIVGPMKEVSDWLTNGIWLDTTGTGL 313
QY 114 GLLFLGLVYSPVITVGLHQSFPPIELEL---FNQG---GSFIFATSMANIAQAACLAFL 167
DB 314 GMGVFGLVSPVVTGLHQSFPPIELELQISAFQNGTGHGDFIFVTASMANVQAAGATFAI 373
QY 168 FFAKSEKLGLAGAGSVASVIGTIPAFVGNLRLRWPFIFIGTAAGIALGALIALFNK 227
DB 374 YFLTOKKMKGLSSSSGVSALLGITTEPALFGLVNLKYRPFPPFCALIGSASAAITGLQV 433

```

```
QY 228 AVALGAAGFLGVVSDIDAPDMVMFLCAVVTETFAAGATAYGLVLYRRNGSIDPDATAP 287
DB 434 AVSLGAGFLGLFUSIRASSIPFVVCCELLISFAIAFTVYGYG-----KTKAVDFVFAEEA 488
QY 288 VPAGTTKAEAEAPAEFNSDST-----IIQAPLTGEATLSSVSDAMFASGKLGSGVAI 340
DB 489 VEBIAEEVQ-EIPEEAASAANKAQVTDVLAAPLAGEAVELTSTVDFVSEAMGKGI 547
QY 341 VPTKGLVSPVSGKIVVAPSPSHAFVAFRKAEDGNSVDILMHIGFTVNLNGTHFNPLKK 400
DB 548 KPSGNTVYAPVDGTQVAFDTGHAYGI--KSDNGA--EILIHIGIDTVSMGKGFQKVQ 603
QY 401 QGDEVKAGELLCFEDIDAIKAAGYEVTTPIVWSNYKKTGPVNTY-GLGETEAGANLLNVA 459
DB 604 ADOKIKKGDVLGTFDSKDAEAGLDNTMTFIVTADYASVETLASSGVAVGDSLLEVK 663
QY 460 K 460
DB 664 K 664

RESULT 3
PTSB_STAXY
ID PTSB_STAXY STANDARD; PRT; 480 AA.
AC P51184;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIIBC COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERWEASE IIIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Staphylococcus xylosoyus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1288;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=DSM 20267 / C2A;
RX MEDLINE=94049686; PubMed=82322209;
RA Wagner E., Goetz F., Brueckner R.;
RT "Cloning and characterization of the scrA gene encoding the sucrose-
RT specific Enzyme II of the phosphotransferase system from
RT Staphylococcus xylosoyus.";
RL Mol. Gen. Genet. 241:33-41(1993).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X69800; CAA49461.1; -
CC InterPro: IPR001996; PTS_EIIB.
CC InterPro: IPR003352; PTS_EIIC.
CC Pfam: PF02378; PTS_EIIB; 1.
CC Pfam: PD001476; PTS_EIIB; 1.
CC ProDom: PD001476; PTS_EIIB; 1.
CC PROSITE: PS01035; PTS_EIIB_CYS; 1.
```

```
KW Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Inner membrane; Phosphorylation.
FT DOMAIN 1 ? EIIIB DOMAIN.
FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 480 AA; 51326 MW; AB4E1D9785D84E47 CRC64;

Query Match 29.7%; Score 701; DB 1; Length 480;
Best Local Similarity 51.4%; Pred. No. 3e-39;
Matches 146; Conservative 49; Mismatches 81; Indels 8; Gaps 4;

QY 1 MAMVPSLVNGYDVAATMAAG-EMPWMSLEGLDVAQAGYOGTVLPVLVYVSWILATIEKPL 59
DB 194 MILVHPMLMSAYDYPKALEAGKEIPHNLFGLLEINQVGVQGLPMLVAVYIILATIEKGL 253
QY 60 HKRLKGTADFLTPVLTLLLTGFTLTAIGPAMRWGDVLAHGLQGLYDFGPGVGLLFG 119
DB 254 RKVIPVLDNLLTPLLAILSTGFTSFVGLTRTGLYWLSDGLTWLYEFGGAIGLIFG 313
QY 120 LVSPVITITGLHOSFPPIELELF----NOGGSPIFATASMANIAQAACLAIVFLAK-SE 174
DB 314 LLVAPIVITCMHHSFIAETQLIADSSSTGSGSFIFPIATMSNIAQAALAAAFIIEK 373
QY 175 KKLGLAGASGVSAVLGITPAIFGVNLRWLPFFIGITGAAGALIALFNKAVAGAA 234
DB 374 KLGKVASAAGVSALLGITPEFAMFVGNLKRYPFIFGAIVGSGISAYIAFFKVAIALGTA 433
QY 235 GLFGVVISIDAPD--MYMFLVCAVVTFFIAFGAAIAYGLVLVRN 276
DB 434 GIPGFISIQNNQNLHYGIAMIIAIFAGVYVALSYRKKYRN 477

RESULT 4
PTSB_PASMU
ID PTSB_PASMU STANDARD; PRT; 474 AA.
AC Q9CJ22;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIIBC COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERWEASE IIIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-SCR).
GN SCRA OR PTSB OR PM1846.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN SEQUENCE FROM N.A.
RP STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC DR EMBL; AE006222; AAK03930.1; -.
CC DR InterPro; IPR001996; PTS_EIIB.
CC DR InterPro; IPR003352; PTS_EIIC.
CC DR Pfam; PF00367; PTS_EIIB; 1.
CC DR Pfam; PF02378; PTS_EIIC; 1.
CC DR PROSITE; PS01035; PTS_EIIB_CYS; FALSE_NEG.
CC DR Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
FT DOMAIN 1 ? EIIB DOMAIN.
FT FT 109 474
FT TRANSMEM 109 129
FT TRANSMEM 158 178
FT TRANSMEM 182 202
FT TRANSMEM 229 249
FT TRANSMEM 264 284
FT TRANSMEM 303 323
FT TRANSMEM 345 365
FT TRANSMEM 376 396
FT TRANSMEM 403 423
FT TRANSMEM 444 464
FT MOD_RES 26 26
FT MOD_RES 324 324
SQ SEQUENCE 474 AA; 49908 MW; D26FAIA059603C01 CRC64;

Query Match 28.7%; Score 678; DB 1; Length 474;
Best Local Similarity 49.5%; Pred. No. 9,7e-38;
Matches 135; Conservative 51; Mismatches 83; Indels 4; Gaps 1;

Qy 1 MAMVPSLVNGYDVAAATMAAGEMPMSLGLDVAQGGVTLVPLVSVWTLATIEKPLH 60
Db 194 MLLVHPALSDGWNVALTLAKNIQVHIFGLEIERVGGVTPVPLVASWVLTATLNLR 253

Qy 61 KRLAGTADFLTPVLTLLTGLTFIAGPAMRWGVDLAHLGLOGLYDFGPGVGLLFL 120
Db 254 KVPFSFLNLTLPALFILTGLTAFTVIGPIGREGAGSLISTGLTWLXYDTLFGVGGAIPT 313

Qy 121 VYSPIVITGLHSPPPLELEFNQ----GGSFIFATASMANIAQGAACLAFFFLAKSEKL 176
Db 314 LYAPVITGMHTFIATVETQLLAEVARTGGTFIFPIRAMSNIAQGAACLAAGVYWKDAKV 373

Qy 177 KGLAGASVSAVLGITEPAIFGVNLRWPPFICGTAAIGTAAIGGALLIATFNKAVLAGAGF 236
Db 374 RGIAPVSGISALLGITEPAIFGVNLRWPPFISAMIGAGISSAVIALFNKAVIALGAAGL 433

Qy 237 LGVVSIDAPDMVFLVCVAVTFEIAFGAAIAYG 269
Db 434 PGIPSIRKPSDLAMVCVGLMISASIAFTLTVLG 466

RESULT 5
ID PTBA_BACSU STANDARD; PRT; 609 AA.
AC P40739; Q45661;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (ETIABC-BGL)
DE (BETA-GLUCOSIDES-PRIMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE
DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (ETI-BGL).
GN BGLP OR N17C.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;

RX MEDLINE=95189730; PubMed=7883710;
RA le Coq D.G., Lindner C., Krueger S., Steinmetz M., Stuelke J.;
RT "New beta-glucoside (bgl) genes in Bacillus subtilis: the bglp gene
RT product has both transport and regulatory functions similar to those
RT of BglF, its Escherichia coli homolog.";
RL J. Bacteriol. 177:1527-1535(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSCIAL;
RX MEDLINE=95219086; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapA loci.";
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE OF 1-182 FROM N.A.
RC STRAIN=168;
RX MEDLINE=96204517; PubMed=8628237;
RA Beloin C., Hirschbein L., le Hegarat F.;
RT "Suppression of the Bgl+ phenotype of a delta hns strain of
RT Escherichia coli by a Bacillus subtilis antiterminator binding
RT site.";
RL Mol. Gen. Genet. 250:761-766(1996).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

FT TRANSMEM 174 194 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 351 371 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 412 432 POTENTIAL.
FT TRANSMEM 452 472 A -> S (IN REF. 3).
FT TRANSMEM 488 508 L -> F (IN REF. 2).
FT TRANSMEM 528 548 E -> G (IN REF. 2).
FT TRANSMEM 568 588 A -> S (IN REF. 2).
FT TRANSMEM 608 628 DG -> HR (IN REF. 2).
FT TRANSMEM 648 668 I -> M (IN REF. 2).
FT TRANSMEM 688 708 G -> S (IN REF. 2).
FT TRANSMEM 728 748 5F630C671D21FBED CRC64;
FT TRANSMEM 768 788 609 AA; 64550 MW; 5F630C671D21FBED CRC64;
FT TRANSMEM 808 828 27.8%; Score 657; DB 1; Length 609;
FT TRANSMEM 848 868 33.4%; Pred. No. 3.1e-36;
FT TRANSMEM 888 908 153; Conservative 84; Mismatches 181; Indels 40; Gaps 8;

QY 13 DVNATWAGEMPWLSFLGLDVAQAGVQGVTLVPLVSVWILATIEKFLHRLKGTADFLIT 72
DB 186 DLTALLGAGK-PI-SPIGLEPVAATSYSTVPIILLSINIASYVEKWKIDRTHASLKLIV 243
QY 73 PVLTLTLLTGLFIAGPAMRWGVDLANGLOGLYDFGGVGLLFGVSPVITGLHQ 132
DB 244 PFTLLIVVPLTLITVGPLGAILGEYLSSGVNVLFDHAGLVAMILLAGTFLSIIMTGMHY 303
QY 133 SFPPIELEFLNCGG-SFIFATASMANIAQAGACLAFFFLAKSEKLLAGAGSVAVLGI 191
DB 304 AFVPIINNIAQNGHDYLLPAMFLANWAGAGASFAVFLSRNKKPKSLALTSITALMGI 363
QY 192 TEPAIFGVNLRWPFPIGIGTAAIGGALIALFNKIKAVALGA-AGFLGVVSDIDAPDMWF 250
DB 364 TEPAMGVNMLKKPPAAALIGAGAGAFYGTGVSAYIVGGNAGLPSVPVFTGPTFIYA 423
QY 251 LVCVVTFEATGAATAYGLYLVRNRSDIDPDATAAPVAGTTKAEAPAESNDSTII 310
DB 424 MGLVTAFAETAAAYLLGFEDVPDSG-----QOPAVHEGSEII 464
QY 311 QAPLTGEATLSVSDAMFASGKLGSGVAIVPTKGLVSPVSGKIVVAPSPGHAFVRTK 370
DB 465 HSPKGEKVALSEKDVGSAGVMGKGFALPEGEVSPVSGVTTIPKTHAIGITS- 523
QY 371 AEDGSNVDILMHIGFTVNLNGHFNPLKQGDVVRKAGELCEFDIDAIAKAGYEVTTPI 430
DB 524 ---DOGAELIHLIGLDTVKLEGOWFTAHIKEGDKVAPGDPVLSFDELEQIKAGYDVITPV 580
QY 431 VVSNYKK--TGPVNTYGLGEIEAGANLLNVAKKEAVPA 466
DB 581 IVTNTDQYSFSPVKEIG-----KVQPEALLA 607

RESULT 6
PTSB_VIBAL
ID PTSB_VIBAL STANDARD; PRT; 479 AA.
AC P22825;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIIB COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERMEASE IIIB COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Vibrio alginolyticus
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
ON NCBI_Taxid=663;
RX MEDLINE=91071601; PubMed=2174811;
RA Blatch G.L., Scholte R.R., Woods D.R.;

```

```

RT "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
RL uptake-encoding region.";
RN Gene 95:17-23(1990).
RP SEQUENCE OF 1-12 FROM N.A.
RX MEDLINE=91285433; PubMed=2060795;
RA Blatch G.L., Woods D.R.;
RT "Nucleotide sequence and analysis of the Vibrio alginolyticus scr
RL repressor-encoding gene (scrR).";
RL Gene 101:45-50(1991).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHIISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M76768; AAA27555.1; -
CC EMBL: M35009; AAA27557.2; -
CC PIR: JQ0781; JQ0781.
CC HSSP: P05053; IIBA.
CC InterPro: IPR001996; PTS_EIIB.
CC InterPro: IPR003352; PTS_EIIC.
CC Pfam: PF00367; PTS_EIIB; 1.
CC Pfam: PF02378; PTS_EIIC; 1.
CC PRODOM: PD001476; PTS_EIIB; 1.
CC PROSITE: PS01035; PTS_EIIB_CYS; 1.
CC KW Phosphotransferase system; Sugar transport; Transferase;
CC Transmembrane; Inner membrane; Phosphorylation.
CC FT DOMAIN 1 ? EIIIB DOMAIN.
CC FT DOMAIN 2 ? 479 EIIIC DOMAIN.
CC FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
CC FT SEQUENCE 479 AA; 49890 MW; FBF906B5170E3EB7 CRC64;

Query Match 27.3%; Score 646; DB 1; Length 479;
Best Local Similarity 46.5%; Pred. No. 1.3e-35;
Matches 128; Conservative 54; Mismatches 87; Indels 6; Gaps 3;

QY 1 MAMVPSLVNGYDVAATMAAGEMPMSLFGDLVAQAGYQGTVPVLPVSVWILATIEKFLH 60
DB 194 MLMVHPLLNGWFGGASVSGNIPVWNILGFEIQKVGQSVLPVLSAFLAKVELGLR 253
QY 61 KRLKGTADFLITPVLTLTLLTGLTFTIAGPAMRWGVDLANGLOGLYDFGGVGLLFLG 120
DB 254 KVTPSVLDNLLTPELLAIFAGLLTFTVVGPFTRDIGLLDGLNWLNTAGVGGAVFGL 313
QY 121 VYSPVITGLHGSFPPIELELF-----NQGSFTIFATASMANIAQAGACLAFLAKSEKL 176
DB 314 IVAPFVITGMHHSFIAIETQLADIATTGCTTFIFPAAAMSNVSGAAALAVGVMSDKKM 373
QY 177 KGLAGASGVSAVLGITEPAIFGVNLRWPFPIGIGTAAIGGALIALFNKIKAVAGAGF 236
DB 374 KGIAIPSGVTGLLGITEPAMFGVNLKLRYPFIAAVCAALSSAFITMNFVKAQALGAAGL 433
QY 237 LGVWSDIDAPDMVNFVCA-VVTFEIAFGAIAAYGL 270
DB 434 PGIISI-TDPKIGYIYAGWVIAFLTAFLVTLVGI 467

```

```

RESULT 7
PTBA_ERWCH
ID PTBA_ERWCH STANDARD; PRT; 631 AA.
AC P26207;
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIBAC COMPONENT (EIIABC-BGL)
DE (BETA-GLUCOSIDES-PERMEASE IIBAC COMPONENT) (PHOSPHOTRANSFERASE
DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).
GN ARBF.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_Taxid=556;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121114; PubMed=1732212;
RA el Hassouni M., Henrissat B., Chippaux M., Barras F.;
RT "Nucleotide sequences of the arb genes, which control beta-glucoside
RT utilization in Erwinia chrysanthemi: comparison with the Escherichia
RT coli bgl operon and evidence for a new beta-glycosylase family
RT including enzymes from eubacteria, archaeobacteria, and humans.";
RL J. Bacteriol. 174:765-777(1992).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81772; AAA24814.1; -
CC PIR; B42603; B42603.
CC HSSP; P20166; 1AX3.
CC InterPro; IPR001127; PTS_EIIA.
CC InterPro; IPR001996; PTS_EIIB.
CC InterPro; IPR003352; PTS_EIIC.
CC Pfam; PF00358; PTS_EIIA.1; 1.
CC Pfam; PF00367; PTS_EIIB.1.
CC Pfam; PF02378; PTS_EIIC.1.
CC ProDom; PD001476; PTS_EIIB; 1.
CC ProDom; PD002243; PTS_EIIA; 1.
CC ProSITE; PS00371; PTS_EIIA.1; 1.
CC ProSITE; PS01035; PTS_EIIB.CVS; 1.
CC Phosphotransferase system; Sugar transport; Transferase;
CC Phosphorylation; Transmembrane; Inner membrane.
CC DOMAIN 1 41
CC ? ? EIIB DOMAIN.
CC DOMAIN 501 631
CC ? ? EIIC DOMAIN.
CC DOMAIN 26 26
CC MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 553 553 PHOSPHORYLATION (BY SIMILARITY).
CC TRANSMEM 50 70
CC TRANSMEM 99 119
CC POTENTIAL.

```

```

FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
FT TRANSMEM 248 268 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 295 315 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT TRANSMEM 358 378 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
SQ SEQUENCE 631 AA; 66984 MM; 7D0BD27A36BFFDFD CRC64;

Query Match 27.1%; Score 640; DB 1; Length 631;
Best Local Similarity 33.8%; Pred. No. 4.2e-35;
Matches 154; Conservative 86; Mismatches 200; Indels 16; Gaps 8;

QY 3 MYPFSLVNGYDVAAATMAAGEMPMSLFLGLDVAQAQYQGTLPVLVLSWILATIEKFLHR 62
DB 184 LVHPSMTAATN---AMQAPDSTLHFLGIPITFTINSSSVIPILFASWSCKLEKPLNRW 240
QY 63 LKGTADFLITPVLTLLTGLFTIAIGPAMRWGDVLAHGLQGLYDFGPGVGLLFLGVY 122
DB 241 LHANIRNFFTPLLCIVISVPLTFLIGPSATWLSQMLAGGVQWLYGLNSLLAGAVMGALW 300
QY 123 SPIVITGLHOSFPPIELFLNQ-GGSFIFATASMANIAQAACLAFLAKSEKLGAG 181
DB 301 QVCVIFGLHGFVPLMLNFSVIGHDTLLPLLVPVAVLGQAGATLGLVLLRTQDLKRKG 360
QY 182 ASGVSALVGTETPAIFGVNLRWPFIFIGTAAIGGALIALEFNKAVALCAAGFLGVVS 241
DB 361 SFSAAIFGTEPAVGVTLPLRPFFIGCGGALGAHVAGYAHHTWYSGFPFISFTQ 420
QY 242 IDAPDMVNLV-CAVYTFITAG-AAIAYGLYLVRRNSIDPDATAAPVPAGTTKABEA 299
DB 421 VIPPTGVDSSWAAVIGTLLAFAPALTSWSFGVPKD---ETQPAADSPA--VLAETQA 475
QY 300 PAEFSNDSTIIQAPLTGEATLSVSDAMFASGLSGVAIVPTKQLVSPVSKIVVAF 359
DB 476 NAGAVRDETILF-SPLAGEVLLLEQVADRTFASGVNMGKGIARPTQGRLYAPVDGTVASLF 534
QY 360 PSGHAFVATKAEKSGNVNIDLMHIGFDTVNLNGHFNPLKQKQDEVKAGELLCFDDAI 419
DB 535 KTHHAIGLASR---GGAELVHIGIDTVLDGRYFFPHVRVGVQVVRQGDLLLEDFGP 590
QY 420 KAAGYEVTTPIVSNYKKTGPVNTYGLGEIEAGANL 455
DB 591 EAAGYDLTTPITVITNSDYRGVPAVSGKVDANAPL 626

```

```

RESULT 8
PTTB_BACSU
ID PTTB_BACSU STANDARD; PRT; 470 AA.
AC P39794; O34771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE)
DE (TREHALOSE-PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC
DE COMPONENT) (EC 2.7.1.69) (EII-TRE).
GN TREP OR TREP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / MABURG;
RX MEDLINE=97074649; PubMed=8917076;
RA Schoeck F., Dahl M.K.;
RT "Analysis of DNA flanking the treA gene of Bacillus subtilis reveals

```

genes encoding a putative specific enzyme IITre and a potential regulator of the trehalose operon.";

Gene 175:59-63(1996).

[2]

SEQUENCE FROM N.A.

STRAIN=AC327;

MEDLINE=97124190; PubMed=8969503;

Yamamoto H., Uchiyama S., Sekiguchi J.;

"Cloning and sequencing of a 40.6 kb segment in the 73 degrees-76

degrees region of the Bacillus subtilis chromosome containing genes

for trehalose metabolism and acetoin utilization.";

Microbiology 142:3057-3065(1996).

[3]

SEQUENCE FROM N.A.

STRAIN=AC327;

MEDLINE=97417486; PubMed=9272861;

Yamamoto H., Uchiyama S., Nugroho F.A., Sekiguchi J.;

"Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region

of the Bacillus subtilis genome reveal genes for a new two-component

system, three spore germination proteins, an iron uptake system and a

general stress response protein.";

Gene 194:191-199(1997).

[4]

SEQUENCE OF 324-470 FROM N.A.

STRAIN=168;

MEDLINE=95379486; PubMed=7651129;

Helfert C., Gotsche S., Dahl M.K.;

"Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by

a phospho-alpha-(1-1)-glucosidase encoded by the treA gene.";

Mol. Microbiol. 16:111-120(1995).

-1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

-TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE

AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY

PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS

PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

THE SUGAR.

-1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOTRANSFERASE + SUGAR =

PROTEIN HISTIDINE + SUGAR PHOSPHATE.

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.

-1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

-1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; 254245; CAA91014.1; -

EMBL; D83967; BAA23409.1; -

EMBL; D86417; BAA22289.1; -

EMBL; X80203; CAA56494.1; -

EMBL; 299108; CAB12609.1; -

HSP; P05053; IIBA.

Subtilist; BG11009; trep.

InterPro; IPR001996; PTS_EIIB.

InterPro; IPR003352; PTS_EIIC.

Pfam; PF00367; PTS_EIIB; 1.

Pfam; PF02378; PTS_EIIC; 1.

ProDom; PD001476; PTS_EIIB; 1.

PROSITE; PS01035; PTS_EIIB_CYS; 1.

Phosphotransferase system; Sugar transport; Transferase;

Transmembrane; Inner membrane; Phosphorylation; Complete proteome.

DOMAIN 1 ?

DOMAIN ?

MOD_RES 26

TRANSMEM 110

TRANSMEM 160

TRANSMEM 183

TRANSMEM 203

TRANSMEM 234

TRANSMEM 254

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

POTENTIAL.

FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 301 321 POTENTIAL.
FT TRANSMEM 326 346 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 403 423 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
FT CONFLICT 140 140 F -> S (IN REF. 1).
FT CONFLICT 363 363 M -> L (IN REF. 1 AND 4).
FT CONFLICT 465 465 A -> G (IN REF. 1 AND 4).
SQ SEQUENCE 470 AA; 49999 MW; 7A741850A2697D53 CRC64;

Query Match 24.1%; Score 570; DB 1; Length 470;

Best Local Similarity 42.8%; Pred. No. 1.2e-30;

Matches 116; Conservative 54; Mismatches 99; Indels 2; Gaps 1;

QY 3 MFVPSLVNGYDVAATMAAGEMPMSLFLGIDVAQAQYQGTVPVSVILATIEKFLHKKR 62
DB 197 LVHPDLLNANGVGAARQSGEIPVNLFLGLEVKYQGVQLPILLASYMLAKIEVFLTKR 256
QY 63 LKGTADFLITPVLLTLLTGFTFIAIGPAMRWGDVLAHGLQGLYDFGPGVGLLFLGLVY 122
DB 257 TPEGIQLLVAVPITLLTGTGFASFIIGPITFAIGNVLTSLISVFGSFAALGGLLYGGFY 316
QY 123 SPIVITGLHQSPFPIELELFNQ--GGSFIFATASMANIAQAACLAFLAVFLAKSEKLGLA 180
DB 317 SALVTGTGHHHTFLAVLDLQIGSKLGTFLWPMALSNIAQGSAAALMMFIVKDEKOKGLS 376
QY 181 GASGVSAVLGITEPAIFGVNLRWRPFPIGIGTAIGGALIALFNKAVAGAAAGFLGVV 240
DB 377 LTSGSIAVLGITEPAIFGVNLRWRPFPIGIGTAIGGALIALFNKAVAGAAAGFLGVV 240
QY 241 SIDAPDMVNLVCAVVTFFIAFGAAIAYGLY 271
DB 437 SIMSQYGAFAIGMAIVLIVPEAGTVAYARE 467

RESULT 9

PTBA_ECOLI

ID PTBA_ECOLI STANDARD; PRT; 625 AA.

AC P08722;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIABC COMPONENT (EIIABC-BGL)

DE (BETA-GLUCOSIDES-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE

ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).

GN BGLF OR BGLC OR BGLS OR B3722.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

[1]

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=88009877; PubMed=3309161;

RA Bramley H.F., Kornberg H.L.;

"Nucleotide sequence of bgIC, the gene specifying enzymebgl of the

PEP:sugar phosphotransferase system in Escherichia coli K12, and

overexpression of the gene product";

RL J. Gen. Microbiol. 133:563-573(1987).

[2]

RN SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=87222180; PubMed=3034860;

RA Schnetz K., Toloczkyi C., Rak B.;

"Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide

sequence, genetic organization, and possible evolutionary

relationship to regulatory components of two Bacillus subtilis

genes.";

RL J. Bacteriol. 169:2579-2590(1987).

[3]

RN


```
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -1- SIMILARITY: 57% IDENTITY WITH THE NEGATIVE REGULATORY PROTEIN OF
CC SACY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03006; AAA22727.1; -.
CC EMBL: X73124; CAA51605.1; -.
CC EMBL: Z99123; CAB15831.1; -.
CC PIR: A39938; A39938.
CC PIR: S39704; S39704.
CC HSP: P05053; IIBA.
CC Subtilist; BG10595; SACP.
CC InterPro: IPR001996; PTS_EIIB.
CC InterPro: IPR003352; PTS_EIIC.
CC Pfam: PF00367; PTS_EIIB; 1.
CC Pfam: PF02378; PTS_EIIC; 1.
CC ProDom: PD001476; PTS_EIIB; 1.
CC ProSITE: PS01035; PTS_EIIB_CYS; 1.
CC Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Phosphorylation; Complete proteome.
FT DOMAIN 1 ? EIIB DOMAIN.
FT MOD_RES 26 460.
FT MOD_RES 309 309.
FT TRANSMEM 113 133.
FT TRANSMEM 148 168.
FT TRANSMEM 178 198.
FT TRANSMEM 202 222.
FT TRANSMEM 247 267.
FT TRANSMEM 288 308.
FT TRANSMEM 328 348.
FT TRANSMEM 358 378.
FT TRANSMEM 386 406.
FT TRANSMEM 407 427.
FT TRANSMEM 429 449.
FT TRANSMEM 460 AA; 45002 MW; 6A303DC042BFE379 CRC64;
SQ SEQUENCE 460 AA; 45002 MW; 6A303DC042BFE379 CRC64;

Query Match 16.5%; Score 390; DB 1; Length 460;
Best Local Similarity 44.6%; Pred. No. 8.7e-19;
Matches 91; Conservative 34; Mismatches 69; Indels 10; Gaps 4;

QY 68 DFLITPVLTLTLGELTFIAIGFAMRWGVGDVLAHGLQGLDFGPGVGLGLVSPDIVI 127
Dd 246 DLLVTEPVTVIVVTCFVAIFAGLPLGRAGSGITVAITVYVDHAGFVAGLFGTYSILVL 305
QY 128 TGLHQSPPTIELEF-NOGGSFIFATASMANIAQGAACLAFFLAKSEKLGAGAGSVS 186
Dd 306 TGVHSHFALEAGLIADIGNKYLPTWSMANVAGGAGLGAFFVMAKAKTKETALPAFS 365
QY 187 AVLGITPEAIFGVNLRWRPFIFIGTGAATGAGLIALFNKAVAGLGAAGFLGVVSDAP- 245
Dd 366 AFLGITPEVIFGVNLRKPFIAAMIGGALGGAVVFTVHVAANAYGLTG-IPMIAAAPP 424
QY 246 ---DMWNFL-----VCVVVTFIAF 262
Dd 425 GFSNLIHYLIGMAIAVSAFIAAF 448

RESULT 15
ID PTTB_ECOLI STANDARD; PRT; 473 AA.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
```

```

FT DOMAIN 1 ?
FT DOMAIN ? 473
FT MOD_RES 29
FT TRANSMEM 111 131
FT TRANSMEM 159 179
FT TRANSMEM 189 209
FT TRANSMEM 226 246
FT TRANSMEM 259 279
FT TRANSMEM 301 321
FT TRANSMEM 341 361
FT TRANSMEM 371 391
FT TRANSMEM 399 423
FT TRANSMEM 441 461
FT CONFLICT 126 127
FT CONFLICT 184 194
FT CONFLICT 187 187
FT CONFLICT 307 307
FT CONFLICT 429 473
FT SEQUENCE 473 AA; 51080 MW; 7437F9822B624944 CRC64;

Query Match 15.5%; Score 367; DB 1; Length 473;
Best Local Similarity 30.1%; Pred. No. 2.9e-17;
Matches 80; Conservative 63; Mismatches 109; Indels 14; Gaps 5;

QY 1 MAMVPSLVNGYDVATMAAGEMP-MWSLFGLDVQAQGYQGTVPVLVYVSWILATIEKFL 59
Db : : | | : | : : : : : : : : : : : : : : : : : : : : : :
195 VTLVSPQLMAY-----LLGQQLPEWDFGFMFSIAKVGYQAQVIPALLAGLGVETRL 249
QY 60 HKRLKGTADFLITPVLTLTLTGLTFTIATGPAMRWGDVLA---HGLOGLYDFGCGPVG 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
250 KRIVPDYLYLVVVPVCSLILAVFLAHLALGPGRMIGDGVAFVAVRHMTGSE---APIGA 306
QY 116 LLFGLVSPIVITGLHQSPPIELF-NQGGSFIPATASMANIAQGAACLAVFFLAKSE.174
Db ||| : ||| : ||| : ||| : : ||| : : : ||| : : : : :
307 ALFGFLYAPLVITGVHOTTLAIDLQIMQSMGCTPWPLIALSNIAQGSAVIGIISSRKH 366
QY 175 KUKGLAGASGVAVLGITEPAIFGVNLRURWPFFIGTAAIGGALIALFNKAVAGAA 234
Db : : : : ||| : ||| : ||| : ||| : : : : : : : : : : : :
367 NEREISVPAAISAWLGVTEPAMYGINLKYRFPMLCAMIGSGLAGLICGLNGVMANGIGVG 426
QY 235 GLFGVVSIDAPDWMVFLVCVVTFEI 260
Db | : ||| : : : : : : : : : : : : : : : : : : : : :
427 GLPGILSIQPSYQWQVFAIAMAIAIII 452
```

Search completed: March 21, 2002, 16:28:49
Job time: 329 sec

THIS PAGE BLANK (USPTO)

Qy	63	LKGTADFLITPVTLTLLTFTTAIGPAMRWGDVLAHGLQGLYDFGPGVGLLFLGVY	1221
Db	242	VPEALDILLTPFLTLMTITGFFAMVVIGPGRFVGDSEISLQTLTYNTTGFSGVLFGGLY	301
Qy	123	SPVITLGHOSFPPIELFLNQGG---SFIFATASMANIAAGAACLAFFLAKSEKLGL	179
Db	302	SLVITGTHHSFHAIEAGLANPAHKNFLLPTWSMANVAGGAALAVYFKTRDKMKSI	361
Qy	180	AGAGSVSAVLGITETPAIFGVNLRLRWPFETIGTAAIGGALIALFNKAVAGAGFLGV	2399
Db	362	AAPASFCLLIGITETPAIFGVNLRKYTRKPIAGALGGALGGYIVFTKVMATVAGVTGIPGI	4211
Qy	240	VSIDAPDMVAFVCAVVYTFITAFGAATAYGLYLVRNRSGIDPDATAAPVPAGTKAEAEA	2999
Db	422	AIVKQSGFLNVIAMI---LAFGAFIIMVLGIRKEEITEEDLNKETVKNKIKVEEVES	477
Qy	300	PAESNSDSTIIQAPLTGEATALSVDSDMPASGKLSGVAIVPTKGQLVSPVSGKIVVAF	359
Db	478	-----VSPVNGKVYLLKNVPDKTFAEGLIGDIGVDPEDEGEVSPIDGTVVHVF	527
Qy	360	PSGHAFVAVRKAEDGSNVVDILMHIGDPTVNLNTHFNPLKKQGDVKYAGELLCEFFIDAI	419
Db	528	ETKHAIAIMKSK---NCVEMLIHIGDITVKMEGNFKSFINDGEEVKKGDKLIQFDLDLV	583
Qy	420	KAAGYEYVTTPIVVSNNYKKTGPVNT 443	
Db	584	KEKAVSPVLTVTNTHEDMGFVNS 607	
RESULT	3		
Qy	Q9KGI9	PRELIMINARY; PRT; 630 AA.	
AC	Q9KG19		
DT	01-OCT-2000 (T-EMBLrel. 15, Created)		
DT	01-OCT-2000 (T-EMBLrel. 15, Last sequence update)		
DE	01-JUN-2001 (T-EMBLrel. 17, Last annotation update)		
DE	PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.		
GN	BH0296		
OS	Bacillus halodurans.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=86665;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN-C-125 / JCM 9153;		
RA	MEDLINE=20512582; PubMed=11058132;		
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,		
RA	Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,		
RA	Horikoshi K.;		
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus		
RL	halodurans and genomic sequence comparison with Bacillus subtilis.";		
RL	Nucleic Acids Res. 28:4317-4331(2000).		
DR	EMBL; AP01508; BAB04015.1; -		
DR	InterPro: IPR001127; PTS_EIIA.		
DR	InterPro: IPR001196; PTS_EIIB.		
DR	InterPro: IPR003352; PTS_EIIC.		
DR	Pfam; PF00358; PTS_EIIA_1; 1.		
DR	Pfam; PF00367; PTS_EIIB; 1.		
DR	Pfam; PF02378; PTS_EIIC; 1.		
DR	ProDom; PD002243; PTS_EIIA; 1.		
DR	PROSITE; PS00371; PTS_EIIA_1; 1.		
DR	PROSITE; PS01035; PTS_EIIB_Cys; 1.		
DR	Complete proteome.		
Qy	SEQUENCE 630 AA; 66978 MW; 6F0218011686ADD5 CRG64;		

```
Query Match      29.4%; Score 695.5; DB 2; Length 630;
Best Local Similarity 32.8%; Pred. No.le-31;
Matches 152; Conservative 108; Mismatches 182; Indels 21; Gaps 5;
```

QY 62 RLKGTADFLITPVLTLLLTGFLTEIFAIGPAMRWGVDLAHLGLOQLYDFGPGVGLLFLGV 121
 Db 237 KIHEAVKVFPTPLILLVIVPVVTLIIILGPIGVVLGNGIASVIOEITFSPVLGAI VAGI 296
 QY 122 YSPIVITGHLQSPPIEL-ELFNQGGSFIFATASMANIAQGAACLAFFFLAKSEKLGKLA 180
 Db 297 WQVLVIFGIHWGIIPIILNLSVRGDEVIKAVAPAVFSQAGALGVMLPTKKNKKKALA 356
 QY 181 GASGSAVIGITPEAIFGVNLRWPFPIGTAAGGALIALFNKAKVALGAAGFLGVV 240
 Db 357 GSTSITAFGITPEAVGVPTPLKGEVFNMAVISAAGVGAIVGHYGSVAVAPAGLLTIP 416
 QY 241 SI---DAPDMVFLVCAVVTFFTAAGAAATAGLYLVRRNGSIDPDATAAPVPAGTTKAEA 297
 Db 417 IFVPEGRGFAVIAIISFVIA-----AVLTIVYGVKDPVDDDTLSNESSENEVKR 471
 QY 298 EAPAEFNSDTITQAPLTGEAIALSSVSAMFASGLGSGVAIVPTKGLQVSPVSGKIVV 357
 Db 472 EDDKKEPSASEEIKSPKGEVVPTEVDHVFSSGAMGKGVAPRKEGRLVAPINGTVTS 531
 QY 358 APTSGHAFVTRKAEDGNSVDILMHGTFDTVNLNGTHFNPLKKGDEVKAGELLCFEDID 417
 Db 532 LFETKKAIGITS---DNGTEIFIRHVIGIDTVQLKGEHFTSFIEQGEVAAGDVLLEFDVE 587
 QY 418 AKAAGYEVTTPVWSNYKTKGPVNTYGLGEIEAGANLLNVAK 460
 Db 588 RITAAGDVITPVLTNAKQFNSVQITDKREVTSDDLIIHVIK 630

RESULT 4

Q9KF90 ID Q9KF90 PRELIMINARY; PRT; 636 AA.
 AC Q9KF90;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PTS SYSTEM, BETA-GLUCOSIDE-SPECIFIC ENZYME II, ABC COMPONENT.
 GN BGLP OR BH0595.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 Horikoshi K.;
 RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.",
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AP001509; BAB04314.1; -
 DR InterPro; IPR001127; PTS_EIIA.
 DR InterPro; IPR001127; PTS_EIIB.
 DR InterPro; IPR001996; PTS_EIIB.
 DR Pfam; PF00358; PTS_EIIB.1;
 DR Pfam; PF00367; PTS_EIIB.1;
 DR Pfam; PF02378; PTS_EIIC.1;
 DR ProDom; PD001476; PTS_EIIB.1;
 DR ProDom; PD002243; PTS_EIIB.1;
 DR ProSITE; PS00371; PTS_EIIB.1;
 DR ProSITE; PS01035; PTS_EIIB_CYS.1;
 KW Complete proteome.
 SQ SEQUENCE 636 AA; 68437 MW; 434C0B12311716F2 CRC64;

Query Match 28.5%; Score 673.5; DB 2; Length 636;
 Best Local Similarity 34.0%; Pred. No. 1.8e-30;
 Matches 163; Conservative 78; Mismatches 193; Indels 45; Gaps 9;

QY 2 AMVFPSLVNGYDYAATMAAGEMPWSLF-----GLDVAQAQYQGTVLPLV 47

Db 183 SLVYPTLV-----VLTEGE-PLYTLFTGCTIFESPWHITFLGIPVILMSYATSVPIIL 234
 QY 48 VSWILATIEKFLKRLKGTADFLITPVLTLLLTGFLTFFTAIGPAMRWGVDLAHLGLOGLY 107
 Db 235 AAYFASKVEARLKKIIPDVVKTELVPPFTLLIVVPLTFVIGPIATWAGOLLGQFTLWVY 294
 QY 108 DFGGPVGGLLFLGLVYSPVITGLHQSPPIEL-ELFNQGGSFIFATASMANIAQGAACLA 166
 Db 295 NLSPIITAGAFGLGFWGVFIFGLHMLIPIAINNVLVQGGSDPVLAMVFAASFAQIGAVAA 354
 QY 167 VFPLAKSEKLGLAGASGVSAVLGTEPAIFGVNLRWPFPIGIGTAAIGGALIALFNI 226
 Db 355 WLKIKQKVKTLVSPAFISGIFGVTEPAIYGVTLPLKRPFIISCIAAAGVGAIIGLFRS 414
 QY 227 KAVALGAAGFLGVSI-----DAPDMYMF-LVCAVVTFFTAAGAAIAYGLYLVRRNGSIDP 281
 Db 415 QGYIIGGLGIFGIPSLHPADGMDAGFWGIVIAVVAVF-LGFILTYLFLGKSGNASDEQ 473
 QY 282 DATAAPVAGTTTAAEAAPAEFSNDSITIQAPLTGEAIALSSVSAMFASGLGSGVAIV 341
 Db 474 TETKAHTSTGTGKEE-----TSSPFNGSVITLSEIKDEAFSSGALGEGIAIE 521
 QY 342 PTRGQLVSPVSGKIVVAFPSGHAFVTRKAEDGNSVDILMHGTFDTVNLNGTHFNPLKKQ 401
 Db 522 PSEGLFSPVSGMVTALYPTTHALGIT---DRGAELLHIIGLDITVOLDGKFFTAHTIQ 577
 QY 402 GDEVKAGELLCFEDIDAIKAAGYEVTTPVWSNYKTKGPVNTYGLGEIEAGANLLNVAK 460
 Db 578 GAQVKGEDLLIEFDIKEIRAAGYAVTTPVIVTNHKVQGLFDTDKQVNAVAGDRLLLETR 636
 RESULT 5
 Q9KLT8 ID Q9KLT8 PRELIMINARY; PRT; 479 AA.
 AC Q9KLT8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIIC COMPONENT.
 GN VCA0653.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 Fraser C.M.;
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.",
 RL Nature 406:477-483(2000).
 DR EMBL; AE004395; AAF96554.1; -
 DR TIGR; VCA0653; -
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIB.
 DR Pfam; PF00367; PTS_EIIB.1;
 DR Pfam; PF02378; PTS_EIIC.1;
 DR ProSITE; PS01035; PTS_EIIB_CYS.1;
 KW Complete proteome.
 SQ SEQUENCE 479 AA; 50292 MW; DC0E129229957E3C CRC64;

Query Match 27.9%; Score 659; DB 2; Length 479;
 Best Local Similarity 49.3%; Pred. No. 8.4e-30;
 Matches 133; Conservative 46; Mismatches 87; Indels 4; Gaps 1;

```
Qy 1 MAMVPSLVNGYDYVAATAAGEMPMWSLFGDLVDAQYQGTVPVLVWVSWILATIERFLH 60
Db 194 MLWHPDLLNGWGSASVSQVTWNLTGFEIEKVGQSVLPVLSAYILAKIENGLR 253
Qy 61 KRLKGTADFLITPVLTLTLGFLTFIAIGPMRWGDLVLAHGLQGLYDFGPGVGLLFLGL 120
Db 254 KIVPSVVDNLTPLMLAIFITGFLTFVVGPLTRDVGFLGDLNWLDSAGFVGGLFGE 313
Qy 121 VYSPIVITGLHOSPPTELELF---NOGGSFIFATASMANIAAGAACLAVERFLAKSEKL 176
Db 314 IYAFVITGMHSHFIAETQLLADIIVTGGTFIPIAAMSNIAGAAALAVGVMTKETKL 373
Qy 177 KGLAGASVSAVLGITEPAIFGNLRLRWPFPIGIGTAAIGGALIALFNKAVAGAAFG 236
Db 374 KGVAIPSGVTALLGITEPAMFVGNLRLRYPIAICGAALASAFITLNVKAQALGAAGL 433
Qy 237 LGVVSIDAPDMVFLVCVAVTFFTAAGAAI 266
Db 434 PGIISINPQIGYINGMAISFVAFAALT 463

RESULT 6
Q99RQ0 PRELIMINARY; PRT; 480 AA.
AC Q99RQ0;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DE PIS SYSTEM, SUCROSE-SPECIFIC IIIB COMPONENT.
GN SCRA OR SA2167.
OS Staphylococcus aureus subsp. aureus N315.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBL_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kihara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43469.1;
KW Complete proteome.
SQ SEQUENCE 480 AA; 51232 MW; D8607F6E0AF2E84 CRC64;

Query Match 27.9%; Score 659; DB 2; Length 480;
Best Local Similarity 48.6%; Pred. No. 8.4e-30;
Matches 139; Conservative 53; Mismatches 82; Indels 12; Gaps 5;

Qy 1 MAMVPSLVNGYDYVAATAAGEMPMWSLFGDLVDAQYQGTVPVLVWVSWILATIERFL 59
Db 194 MLWHPDLLNGWGSASVSQVTWNLTGFEIEKVGQSVLPVLSAYILAKIENGLR 253
Qy 60 HKRLKGTADFLITPVLTLTLGFLTFIAIGPMRWGDLVLAHGLQGLYDFGPGVGLLFLG 119
Db 254 KIVPSVVDNLTPLMLAIFITGFLTFVVGPLTRDVGFLGDLNWLDSAGFVGGLFGE 313
Qy 120 LVYSPIVITGLHOSPPTELELF---NOGGSFIFATASMANIAAGAACLAVERFLAK-SE 174
Db 314 IYAFVITGMHSHFIAETQLLADIIVTGGTFIPIAAMSNIAGAAALAVGVMTKETKL 373
Qy 175 KGLAGASVSAVLGITEPAIFGNLRLRWPFPIGIGTAAIGGALIALFNKAVAGAA 234
Db 374 KGVAIPSGVTALLGITEPAMFVGNLRLRYPIAICGAALASAFITLNVKAQALGAAGL 433
Qy 235 LGVGVVSIDA--PDMWMLVCVAVTFFTAAGAAIAYGLVLRNGS 278
Db 434 PGIISINPQIGYINGMAISFVAFAALT 463
```

```
Db 434 GLPGFISINPVHAGWHLHFVGMTISFII----AITVTILSKRKN 475

RESULT 7
Q9KJ80 PRELIMINARY; PRT; 644 AA.
AC Q9KJ80;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE BETA-GLUCOSIDE-SPECIFIC EII PERMEASE.
GN BGLP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBL_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RA Cote C.K., Cvitkovitch D., Bleiweis A.S., Honeyman A.L.;
RX MEDLINE=20340959; PubMed=10878120;
RT "A novel beta-glucoside-specific PTS locus from Streptococcus mutans
RT that is not inhibited by glucose."
RL Microbiology 146:1555-1563(2000).
DR EMBL; AF206272; AAF89975.1;
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
DR PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.
SQ SEQUENCE 644 AA; 69282 MW; 723B7FBDD2794EB CRC64;

Query Match 27.9%; Score 658.5; DB 2; Length 644;
Best Local Similarity 33.8%; Pred. No. 1.3e-29;
Matches 160; Conservative 89; Mismatches 196; Indels 29; Gaps 12;

Qy 2 AMVFPSSLVNGYDYVAATAAGEMPMWSLFGDLVDAQYQGTVPVLVWVSWIL 52
Db 185 ALVYPNIAASFTVKHPLTYLTFQGTLLVSPISYVTFPIIPFAPSSYLQTLVPIVVAIWAG 244
Qy 53 ATTEKFLHKLKGTADFLITPVLTLTLGFLTFIAIGPMRWGDLVLAHGLQGLYDFGGP 112
Db 245 SKIETFEKKIIPDVVKVVFVVFVFFLITLIVPLSLFVLGPMVMSWASDLVGAIFTGIYGNPV 304
Qy 113 VGLLGLVYSPIVITGLHOSFPPIE--LELFNOGGSFIFATASMANIAAGAACLAVERFLA 171
Db 305 IYGVILGAMWQVLVWFLGHLWGLVPLALELQKGGVILVATIAIC-FAQAGSLINIMRT 363
Qy 172 KSKELKGLAGASVSAVLGITEPAIFGNLRLRWPFPIGIGTAAIGGALIALFNKAV 231
Db 364 NENKROLSPATISALFGLTEPAIYGITLPMRVPFITMTCVSGAISGAYLALFNKQVM 423
Qy 232 GAAGFLGVVSIDAPD---VMVMLVCVAVTFFTAAGRAIAYGL-VLVRNGSIDPDATAA 286
Db 424 GGMGLFAIPSFIDPKNSMILIHFIATAMFVLGVITQFIKIPYLYGETSTSDSDVDDK 483
Qy 287 PVFAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVDSAMFASGKLGSGVAIVPTKQ 346
Db 484 EEPVKELK-----EIKQE--IISPLICKVVKLENVDPDEFASGAMGKGIADPDGDI 534
Qy 347 LVSPVSGKIVVAFPSGHAFVARTKAEDGSNVDILMHIGTDTVNLNGHFNPLKKQGEVK 406
Db 535 VVAPTKEVTLVPTKHAVALRT--ENGA--EILIHGMDTVSLAGKGRKSFVEVGSHVE 590
Qy 407 AGELLCEFDIDAKAAGYEYVTPVIVSNYKTKTPVNTYGLGEIEAGANLLNVAK 460
Db 591 AGQTLLFEDVNAKAAGLPVITPVIVTNSQDFEDVLVTQERTVEAGNLLTAVK 644
```

```
RESULT 8
Q99X04
ID Q99X04 PRELIMINARY; PRT; 674 AA.
AC Q99X04;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE PUTATIVE PTS SYSTEM ENZYME II.
GN SPY2097.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes.*;
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006630; AAK34748.1; -.
KW Complete proteome.
SQ SEQUENCE 674 AA; 72490 MW; 4BAA26B40E784878 CRC64;

Query Match 27.4%; Score 647.5; DB 2; Length 674;
Best Local Similarity 34.4%; Pred. No. 5.8e-29;
Matches 155; Conservative 86; Mismatches 178; Indels 31; Gaps 11;

QY 1 MAMVPSLVNGYDVAAATMAA--GEMPMWSLGLDVAQAQYQGTVPVLVSVWILATIEKFLH 58
DB 212 ICLVSPQLNAYAVAGTAAETAKNWNWDFGFTINRIGYQAQVIPALLAGLSLAYLEIF 271
QY 59 LHKRGKTADFLITPVLLTLLTGLFTFIAGPAMRWGVDVLAHGLQGLYDFG--GPVG-- 114
DB 272 WRKRIPVSMVFVPLSLIPALIAHTVLGP----IGWTIGKISFVVLAGLTGPVKWL 327
QY 115 -GLLGLVSPVITGLHOSFPPIELELFNOGG---SFIFATASMANIAQGAACLAFFVL 170
DB 328 FGAIFGALYAPLVTGLHMTNAIDTQLIADTATRTTGLWPMIALSNIAQGSVAFFAYLM 387
QY 171 AKSEKLG-LAGASGVSALVTEPAIFGVNLRWRPFIFIGTAAIGGALLALFNKAV 229
DB 388 NRHEEREISUPAASATVGLVTEPAIFGVNLRWRPFIFIGTAAIGGALLALFNKAV 447
QY 230 ALGAAGFLGWSIDAPDMVFLVC----AVVTFFIAFGAAIAYGLVLRNGSIDPDATA 285
DB 448 STGVGLPGFMALNVKYMIPFFICMAVAIVVPMFLTFFPKSHMTKTEDEAKL-PETPV 506
QY 286 APVPAGTTKAEAPAEAFSNDSTI-IOAPLTGEAIALSSVDAMFASGKLGSGVAIVPTK 344
DB 507 SDAPVAT-----APHKTMQGTVITLSTPLTGEVKALSEAVDPVPAQGVMGQALLQPT 560
QY 345 QGLVSPVSKIVVAPPSGHAFVTRKAEDGNSVDILMHIGFTDVTNLTGTHFNPLKQGB 404
DB 561 GVLVAPCDAEVSLFPTKHAICLVTT---EGLELHMIGMDVTNLDGGQFEALVRQGGQ 616
QY 405 VRAGELLCFDFDAIKAAQGYEVTTPVWSN 434
DB 617 VRAGQTLQFDIAAISEAGYATETPLVVTN 646

RESULT 9
Q99Y91
ID Q99Y91 PRELIMINARY; PRT; 620 AA.
AC Q99Y91;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE PUTATIVE SUCROSE-SPECIFIC PTS PERMEASE, ENZYME II.
GN SPY2097.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes.*;
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006630; AAK34748.1; -.
KW Complete proteome.
SQ SEQUENCE 674 AA; 72490 MW; 4BAA26B40E784878 CRC64;

Query Match 26.7%; Score 632; DB 2; Length 620;
Best Local Similarity 35.6%; Pred. No. 3.9e-28;
Matches 160; Conservative 66; Mismatches 202; Indels 22; Gaps 6;

QY 1 MAMVPSLVNGYDVAAATMAAAGEMPMWSLGLDVAQAQYQGTVPVLVSVWILATIEKFLH 60
DB 189 LMLVSNELFNWVA---SGGDKVPLTFFGF-VPVVGYQGTVPVPAFFVGLVGAKLEKWLH 244
QY 61 KRLKGTADFLITPVLLTLLTGLFTFIAGPAMRWGVDVLAHGLQGLYDFGPGVGLLFL 120
DB 245 KKVPEALDLLVTFPLTFATIMSTGLFVIGPVPHSLNENLVLAGTQAVLHLPLFGIAGLIVG 304
QY 121 VYSPVITGLHOSFPPIELELF-NOGGSFIFATASMANIAQGAACLAFFLAKSEKLG 179
DB 305 IQGLIVVTGIHIFNLEAQLIANTGKDPFNAYLTAAATAAGATLAVAVTKSTKLKGL 364
QY 180 AGASGVSALVTEPAIFGVNLRWRPFIFIGTAAIGGALLALFNKAVKAAAGFLV 239
DB 365 APSTLSALLGITEPAIFGVNLRWRPFIFIGTAAIGGALLALFNKAVKAAAGFLV 424
QY 240 VSIDAPDMVFLVCAYVTFFIAFGAAIAYGLVLRNGSIDPDATAAPVPAGTTKAEAE 299
DB 425 LLYNLQQLQYLVLTMLVGLGVAFAIATWGY-----QDRETLPLPAVEDQTDQ 474
QY 300 PAFSNDSTIIQAPLTGEAIALSSVDAMFASGKLGSGVAIVPTKQGLVSPVSGKIVVAF 359
DB 475 PALAEE---TLYSPLNGTVVDLSAVSDPVFSSGAMGGGLAIPKPEDNTLYSPVDGKVEIF 531
QY 360 PSCHAFVTRKAEDGNSVDILMHIGFTDVTNLTGTHFNPLKQGBDEVKAGELLCFDFDAI 419
DB 532 ETGHATITS-----SOGAEVLLHIGIDTSMAGDGFESLVAVGQAVKKGDLGHDFDSKI 587
QY 420 KAAQYEVTTPIVVSNNKKTGPVNTYGLGEI 449
DB 588 AEAGLDTTMMIVSNIADYQSDVILAAGHV 617

RESULT 10
Q48408
ID Q48408 PRELIMINARY; PRT; 621 AA.
AC Q48408;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CELLOBIOSE-SPECIFIC PTS PERMEASE.
GN KLEBIELLA.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
ON NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P2;
```



```

RX MEDLINE=97176370; PubMed=9023916;
RA Lai X., Davis F.C., Hespell R.B., Ingram L.O.;
RT "Cloning of cellobiose phosphoenolpyruvate-dependent
RT phosphotransferase genes: functional expression in recombinant
RT Escherichia coli and identification of a putative binding region for
RT disaccharides";
RL Appl. Environ. Microbiol. 63:355-363(1997).
DR EMBL; U61727; AAB51563.1; -
DR HSP; P20166; IGRP.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIB; 1.
DR ProSITE; PS00371; PTS_EIIA_1; 1.
DR ProSITE; PS01035; PTS_EIIB_CYS; 1.
SQ SEQUENCE 621 AA; 65492 MW; DF1DE2A2A7B81339 CRC64;

Query Match 25.98; Score 612.5; DB 2; Length 621;
Best Local Similarity 32.48; Pred. No. 4.9e-27;
Matches 154; Conservative 89; Mismatches 181; Indels 51; Gaps 11;

QY 2 AMVPSLVNGYDVAAATMAAGEMPWMSLFLDVAQAGYGTGTVLPVLVSVWILATIEKFLHK 61
DB 182 ALTHPLMQAD--ASMAPGAATEYFL-GIPVTINTSSVIPILASWSCWIEKRSNA 238
QY 62 RLKGTADFLTPVLTLLTGLTFIATGPMRWYGDVLAHGLQGLYDFGGVGLLGLV 121
DB 239 ILPSSMKNFPTAICLAVVVPVLTLLIIPVATLWSQLLANGYQLIYQVAPWLAGAANGAL 298
QY 122 YSPIVITGLHOSFPPIEL-ELFNOGGSFIFATASMANIAQAACLAFFFLAKSEKLGKLA 180
DB 299 WQCVIFGLHGLIPLMINNLAVLGHDSMPMLLPVAVMGQVGAALGFLTRDARQKVL 358
QY 181 GASGSAVLGITETPAIFGVNLRWRPFIFIGITAAIGGALIALFNKAVALGAAGFLGV 240
DB 359 GSAYSAGIFGVTEPAIYGLNPLRRPFICGVYPIGAWVG-FSDSHVSTYSFGFNII 417
QY 241 S-----IDAPDMVFLVCAVVTFFIAGAAIAYGLVLRNRSIDPDATAAPYAG 291
DB 418 TLAQIMPEGIDAT-----VMGGAAGMEASLI-----ACVLTAVAG 454
QY 292 TTKAEE-----APAEFNDSTIIQAPLTGEATALSVDAMFASCKLGSVAIVPTKG 345
DB 455 LPRSSAQAAVVPASV-ND---ILAPMTGSLVALDQVDPSTFASGLLGQGVAIIPSVG 510
QY 346 QLVSPVSGKIVVAPPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTHENPLKKGDEV 405
DB 511 KVIAFFSSEVASIPQTKHAIGLLS-----DSGIELLIHVGDITVKLDCAPTAHVKEDKI 566
QY 406 KAGELCEFDIDATAKAGYEVTTPIVVSNYKTKGPVNTYTLGTEIEAGANLLNVAK 460
DB 567 KAGDLLLEFDRLQALDLDAGYDLATPLIITSNSDDFTLDMWSASAVDAGQPLLSVSR 621

RESULT 11
Q9L461 PRELIMINARY; PRT; 577 AA.
AC Q9L461;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PERMEASE.
GN BGLP.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=B21;
RX MEDLINE=20263773; PubMed=10802183;
RA Marasco R., Salatiello I., De Felice M., Sacco M.;
RT "A physical and functional analysis of the newly identified bglGPT
RT operon of Lactobacillus plantarum.";
RL FEMS Microbiol. Lett. 186:269-273(2000).
DR EMBL; AJ250202; CAB71150.1; -
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD002243; PTS_EIIB; 1.
DR ProSITE; PS00371; PTS_EIIA_1; 1.
SQ SEQUENCE 577 AA; 60877 MW; E8C877050103B5FC CRC64;

Query Match 25.74; Score 607; DB 2; Length 577;
Best Local Similarity 32.68; Pred. No. 9.2e-27;
Matches 147; Conservative 81; Mismatches 133; Indels 90; Gaps 12;

QY 2 AMVPSLVNGYDVAAATMAAGEMPWMSLFLDVAQAGYGTGTVLPVLVSVWILATIEKFLHK 61
DB 188 ALLYPTLTVMTNSTTL-----HFFGIPVPTTSTVPIILLAVWVLSYLEPVLDK 239
QY 62 RLKGTADFLTPVLTLLTGLTFIATGPMRWYGDVLAHGLQGLYDFGGVPGV-GLLFLGL 120
DB 240 LFPAAIRNIETPLSLIIMVPLTLLVDGPIGGLINGLASGVMAIYNF-MPIGAGVINGA 298
QY 121 VYSPIVITGLHOSFPPIEL-ELFNOGGSFIFATASMANIAQAACLAFFFLAKSEKLGK 179
DB 299 FQWVFVIFGVHWTVPVLMNNAIKMGYDPLLPILSAVLSOAGALAVFLKSRQKKKAL 358
QY 180 AGASGSAVLGITETPAIFGVNLRWRPFIFIGITAAIGGALIALFNKAVALGAAG----- 235
DB 359 AGSFVTALFGITEPTIYGVTLKLRPFYCAVVGAGLGGAIL-----GAAGTHAS 408
QY 236 -----FLG-----VVSIDAPDMVFLVCAVVTFFIAGAAIAYGLVLRNG 277
DB 409 SFTLPSSLAVPTFLGHGFMGEVIGL-----IVAFGLGAILTYFFGFA----- 450
QY 278 SIDPDATAAPVPAGTTKAEAEAPAEFS-----NDSTIIQAPLTGEATALSVDAMFASG 332
DB 451 -----RQQADYSVVSOSDNTVM-APVEGTIPLTSVHDEFASE 490
QY 333 KLGSGVAIVPTKGOLVSPVSGKIVVAPPSGHAFVTRKAEDGSNVDILMHIGFTVNLNG 392
DB 491 AMKGKLAIVPNKGVTPAVDGTITAVYPTGHAIGITA-----NSGAELIHIGINTVQLNG 546
QY 393 THFNPLAKQDEVKAGELCEFDIDATAKAG 423
DB 547 QYFETMYKQNVKRGDLTKFDVKIKTAG 577

RESULT 12
Q46129 PRELIMINARY; PRT; 616 AA.
AC Q46129;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PTS-DEPENDENT ENZYME II.
GN ABGF.
OS Clostridium longisporum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1523;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6405;
RX MEDLINE=98151780; PubMed=9491080;

```


OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A. 9153;
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AP001514; BAB05935.1; -
 DR InterPro: IPR001996; PTS_EIIB.
 DR InterPro: IPR003352; PTS_EIIC.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
 KW Complete proteome.
 SQ SEQUENCE 470 AA; 50185 MW; 3CE67B1E9650F5B9 CRC64;

Query Match 24.1%; Score 569.5; DB 2; Length 470;
 Best Local Similarity 42.9%; Pred. No. 9.3e-25;
 Matches 112; Conservative 55; Mismatches 93; Indels 1; Gaps 1;

QY 1 MAMVEPSLVNGYDVAAATMAAGEMPMSLFGLDVAQAGYQGTVPVLVLSWILATIEKFLH 60
 DB 195 LMLVHPDLLNANGVQQAQLEGEITWNLFGLTIEQVGTQVQLFSLWSILAKIEIFLR 254
 QY 61 KRLKGTADFLITPVLTLTLTGLFTTFAIGPAMRWGDLVLAHGLQGLYDFGPGVGLLFLG 120
 DB 255 KRVPDSIQLLVAVPALLITGFIATAAIGPTTFTIGNITNVFTSIFAAPVLVGGFLYGL 314
 QY 121 VYSPIVITGLHQSFPPIELELFNQ-GGSFIFATASMANIAAGAACLAFLVFLAKSEKLG 179
 DB 315 IYAPLVVTGMHHTFLAVDLQIGTFLPILVLSNIAQGSAAALAMFATRDEKLG 374
 QY 180 AGAGSVAVLGITEPAIFGVNLRWPFPGTGAATAGGALIALFNKAKVALGAAGFLGV 239
 DB 375 SLSSAVSAYLGITEPAMFGVNIRKFFVCALISAAIGGAFITVNGVLANSIGVGLPGI 434
 QY 240 VSDAPDMVFLVCVVTFPI 260
 DB 435 FSIQAGFWGVFFIGMVIATIL 455

RESULT 15
 Q9A0X4 PRELIMINARY; PRT; 620 AA.
 AC Q9A0X4;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BETA-GLUCOSIDE PERMEASE IIABC COMPONENT.
 GN SPY0572.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF370;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savić D.J., Savić G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL; AE006513; AAK33557.1; -
 KW Complete proteome.
 SQ SEQUENCE 620 AA; 66148 MW; 1AF6872CFDD7C7D6 CRC64;

Query Match 23.9%; Score 564.5; DB 2; Length 620;
 Best Local Similarity 30.8%; Pred. No. 2.5e-24;
 Matches 135; Conservative 93; Mismatches 179; Indels 31; Gaps 9;

QY 3 MVFPSLVNGYDVAAATMAAGEMPMSLFGLDVAQAGYQGTVPVLVLSWILATIEKFLH 62
 DB 182 LLHPAFV-----AMVAEGKPL-TLEGAPVTPASYSSVIPILMVMYLMQYIEKWNRL 233
 QY 63 LKGTADFLITPVLTLTLTGLFTTFAIGPAMRWGDLVLAHGLQGLYDFGPGVGLLFLGV 122
 DB 234 VPSVMKSFLOPTLLIISGFLALVVVGPLGVIIGGSLNTMLAIYHVAPWLAAILGAIM 293
 QY 123 SPIVITGLHQSFPPIELELFNQGSFIFATASMA-NIAQGAACLAFLVFLAKSEKLG 181
 DB 294 PLVMTGMHWAFAFIFLAASVATPDVILPAMLASNLAAQGAASLAVAFKTKQKOTRQ 353
 QY 182 ASGVSAYL-GITEPAIFGVNLRWPFPGTGAATAGGALIALFNKAKVALGAAGFLGV 240
 DB 354 AAGISALLAGITEPALLYGVTLKFKKPLVYAMISGLVGAFIGFVNIASYTFVWPSIIG 413
 QY 241 SIDAP-----DMVMFLVCVVTFPIAAGAAIAYGLYLVRNGSIDPDATAAPVPA 296
 DB 414 QYINPSSGANFTNALIAGTATIVLAFSLTWFMG-----IDEE---SPKQVSVAADM 461
 QY 297 AEAPAEFSNDSTIIQAPLTGEAIALSVSDAMFASGKLGSCVAIVPTKGQLVSPVSGKI 356
 DB 462 SQVSGSLSTKQT-LYAPMTGEMFLSEVPDETFFSKLLGEGFALPSEGEVYAPFDGEVI 520
 QY 357 VAFPSGHAFVAVRTKAEDGSNVQIILMHIGFDTVNLNGTHFNPLKKQGVKAGELLCEFDI 416
 DB 521 TFPPTKHAVALK----NTRGVEVLIHVGIDTVLKGQFQVSVGDVVRKGQALLKMDI 576
 QY 417 DAKAAQYEVTTPIVNSN 434
 DB 577 DFITSKGYSLISPVVVTN 594

Search completed: March 21, 2002, 16:28:26
 Job time: 336 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 07:02:16 ; Search time 66.97 Seconds
(without alignments)
5163.982 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 cttcatgcatctgcgcgtt.....gttgaaaccttgagtgttgcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 15 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	109.6	7.2	3615	1 US-08-920-812-17	Sequence 17, Appl
C 2	109.6	7.2	3615	1 US-08-920-827-17	Sequence 17, Appl
C 3	109.6	7.2	3615	1 US-08-921-177-17	Sequence 17, Appl
C 4	109.6	7.2	3615	1 US-08-362-577C-17	Sequence 17, Appl
C 5	109.6	7.2	3615	2 US-08-920-828-17	Sequence 17, Appl
C 6	97.8	6.4	465	2 US-08-673-190A-3	Sequence 3, Appl
C 7	71.6	4.7	357	2 US-08-673-190A-6	Sequence 6, Appl
C 8	43.2	2.8	7218	1 US-08-232-463-14	Sequence 14, Appl
C 9	37.2	2.4	4403765	4 US-09-103-840A-2	Sequence 2, Appl
C 10	34.6	2.3	46899	1 US-08-471-119A-1	Sequence 1, Appl
C 11	32.8	2.1	12912	2 US-08-460-751-1	Sequence 1, Appl
C 12	32.8	2.1	14060	3 US-08-658-136-4	Sequence 4, Appl
C 13	32.6	2.1	477	4 US-09-135-994-1	Sequence 1, Appl
C 14	32.6	2.1	80161	3 US-09-036-987A-1	Sequence 1, Appl
C 15	32.6	2.1	80161	4 US-09-370-700-1	Sequence 1, Appl
C 16	32.4	2.1	292	4 US-09-117-121-29	Sequence 29, Appl
C 17	32.2	2.1	7218	1 US-08-232-463-14	Sequence 14, Appl
C 18	32	2.1	1008	4 US-09-198-955A-9	Sequence 9, Appl
C 19	32	2.1	1901	1 US-08-153-848-43	Sequence 43, Appl
C 20	32	2.1	1901	3 US-09-299-843A-43	Sequence 43, Appl
C 21	32	2.1	1901	5 PCT-US93-11153-43	Sequence 43, Appl
C 22	32	2.1	2232	1 US-08-241-766-12	Sequence 12, Appl
C 23	32	2.1	2453	5 PCT-US95-07180-1	Sequence 1, Appl
C 24	32	2.1	3120	1 US-08-491-146-1	Sequence 1, Appl
C 25	32	2.1	3120	1 US-08-241-766-11	Sequence 11, Appl
C 26	32	2.1	3120	1 US-08-234-011-1	Sequence 1, Appl
C 27	32	2.1	3120	2 US-08-701-062A-1	Sequence 1, Appl

28 32 2.1 3533 4 US-08-985-916-15 Sequence 15, Appl
C 29 32 2.1 5163 3 US-08-700-651-1 Sequence 1, Appl
C 30 32 2.1 5163 3 US-08-928-361B-4 Sequence 4, Appl
C 31 32 2.1 5318 3 US-08-700-651-2 Sequence 2, Appl
C 32 32 2.1 5318 3 US-08-928-361B-3 Sequence 3, Appl
C 33 32 2.1 10348 2 US-08-457-273B-41 Sequence 41, Appl
C 34 32 2.1 10348 3 US-08-556-419-13 Sequence 13, Appl
C 35 32 2.1 10348 4 US-09-041-886-14 Sequence 14, Appl
C 36 32 2.1 10366 1 US-08-246-982A-5 Sequence 5, Appl
C 37 32 2.1 10366 1 US-08-453-265-5 Sequence 3, Appl
C 38 31.8 2.1 1227 3 US-09-074-912-3 Sequence 2, Appl
C 39 31.6 2.1 5511 3 US-08-928-361B-2 Sequence 2, Appl
C 40 31.6 2.1 7334 3 US-08-928-361B-1 Sequence 1, Appl
C 41 31.6 2.1 7655 1 US-08-619-554-1 Sequence 1, Appl
C 42 31.2 2.0 342 4 US-08-905-223-196 Sequence 196, App
C 43 31.2 2.0 818 4 US-08-998-416-429 Sequence 429, App
C 44 31 2.0 34303 2 US-08-735-609-4 Sequence 4, Appl
C 45 31 2.0 34303 2 US-08-735-609-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-920-812-17/c
; Sequence 17, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
US-08-920-812-17

Db 2435 TATCAACCTGGATAGCGAACTGATTGCGCGGTTATCTGGCTTTATCAGCGGTTTC 2376
QY 432 gtccagtcgagtgctcttcggtctggtctactcaacaatcgtcactcgtcgtcgc 491
Db 2375 CTGCATTTTCGGGCGCGGTAATGGCGGCTTCTGCAAAATCTTCGTCATGTTGCGACTGC 2316
QY 492 accagtcctcccgcaaatgagtcggagctgtttaaccagagtggtgactcctcattctcg 551
Db 2315 ACTGGGGCTGGTGGCGGTGTGTATCAATAACTTCACCTGTGGGCTACGACACCATGA 2256
QY 552 caacggcatctatg---gctaataatgccaggggtgctgcatgtttggcagtggtttcc 608
Db 2255 TCCCGCTGTAATGCCCGCATATATGGCGCAGTTCGGCGCGCGCTCGCGCTTCTCCTCT 2196
QY 609 tggcgaagagtgaaagctcaaggccctgcaagggtcttcaggtgtctccgtgttctctg 568
Db 2195 GCGAAGCGCATGCGCAAAAAAAGTGGTGGCGGATACAGCGCGTGTGACGAGTCTGTTTG 2136
QY 669 gtattacggagcctgcgtcttcctcggtgtgaaccttcgctgctgctgctgttcttcacg 728
Db 2135 GTATCACCAACCGAGGTATATGGGTGTCACCTGCGGCTAAGTACCCCTTTGTTATCG 2076
QY 729 gtatgggtaccgcagctatcggtggcgctttgattgcactctttaaatacaaggcagttg 788
Db 2075 CCTGTATCATGCTGGGCTTTGGGGCCACCATATTGCTACGCGCAACGAAAGTCTACT 2016
QY 789 cgttgggcgtcaggtttcttgggtgtgtttcttctattgatcctcaagatatggtcagt 848
Db 2015 CCTTTGGTTTGCAGTATTTTACCTTTCATGCAAAACCATCCCGTCAACGGGAATTGATT 1956
QY 849 tcttgggtgtcaggtgttacctcttcctcctcctcctcctcctcctcctcctcctcctc 908
Db 1955 TCACCGTCTGGGCCACCGTTATGGCGGTGTCATGTCCTATCGGTCGTCGTCGTCGTCG 1896
QY 909 ttacttggctgcgcgaacgcagcagcattgatccagatgcaacccgctcctcagtgctg 968
Db 1895 CGGTGATGCTTTCATTTTCATCCCGTAAACGTCAGCGCACGCGAGGTGCCCC----- 1842
QY 969 caggaaacacaaagccgaagcagaagaccgcgcagaattttcaaacgattccaccatca 1028
Db 1843 -----GAAAGAAACACCAAGAGTATTATACACCATGAGCAGGCGGTA 1798
QY 1029 tccaggcacccttaccggtgaagctattgactgaagcagcagcagcagcagcagcagcagc 1088
Db 1797 TCTGTTACCGATGACGGGAGAGATGTCGCTCATTCACGTCGCTGATACACGCTTTG 1738
QY 1089 ccagcgaaagcttgctggcggttgccatcgtcccaacaaaggcgagtgagttcttc 1148
Db 1737 CCAGTGGCCTGTTGGTAAAGGTATTGCCATTCGCCCTCGGTGTTGTAAGTCCGTTCTC 1678
QY 1149 cggtagtggaagattgtggtggcattcccatctggtccatcgttcttcgagttcgcacca 1208
Db 1677 CGGTTGCGGGTCCGAATGCTGCTGTTGTCGTCACATTTACA-----CGCCATTGGCATTTG 1624
QY 1209 aggtcagagatgttccaaatggatattcttgaatcacaattggttttcgacagtaaaccc 1268
Db 1623 AGTCAGATAGG-----TGTGGAGATCCTGATTTCATGTCGGTATFCGACACCGTAAAC 1570
QY 1269 tcaacggcagcagcactttaaccgcgtgaagaagcagggcgatgaagtcaaacgagggagc 1328
Db 1569 TGGACGGCAAAATCTTTTCGGCTCAGGTCAAGTGGGTGACAAAGGTCAATACAGGGATC 1510
QY 1329 tgcgtgtgaatcgatattgtagccattaaagctgacggttatgaggttaacaaagccga 1388
Db 1509 GGCTGATTTCTTTTATATCCCTGCTATTTCGCGAGGCGCGGATTTGATCTGACGACGCCG 1450
QY 1389 ttggtgttcgaattacaagaaacggacctgttaaacacttacggttttggcggaatttg 1448
Db 1449 TATTAATAGTAAAGCATGATTTTACGGACGTATTATACCCACGCGCACGCGCAGATAA 1390
QY 1449 aagcgaggaaccaacctgtcacaagtc 1474

Db 1389 GCGCAGGTGAACCGCTGTTTATCCATC 1364
RESULT 3
US-08-921-177-17/c
; Sequence 17, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical isolate EC-24
; US-08-921-177-17

Query Match 7.2%; Score 109.6; DB 1; Length 3615;
Best Local Similarity 46.3%; Pred. No. 2.2e-23;
Matches 540; Conservative 0; Mismatches 589; Indels 37; Gaps 4;

QY 312 tcaactcagtcgctgagctgtgctgcacggattccttaccattcgcgcatcgccag 371
Db 2495 TCACACCATTTGCTATGCTGATGGTTATCACACCGCTCACCTTCTGCTGGTGGGCGCGC 2436
QY 372 caatgcgctgggtggtggatgctgctggcagcaggtctacagggactttatgattcggctg 431
Db 2435 TATCAACCTGGTAAAGCAACTGATTGCGCGGTTATCTGCTGGCTTTATCAGCGGTTTC 2376
QY 432 gtccagtcgagtgctcgtcttcggtctgtactcaccacacgtcactcactcgtcgtcgc 491
Db 2375 CTGCATTTTCGGGCGCGGTAATGGGCGGCTTCTGGCAAAATCTTCGTCATGTTGCGACTGC 2316
QY 492 accagtcctcccgcaaatgagtcggagctgtttaaccagagtggtgactcctcattctcg 551
Db 2315 ACTGGGGCTGGTGGCGGTGTGTATCAATAACTTCACCTGTGGGCTACGACACCATGA 2256


```

QY 729 gtagcgtaccgcagctatcggtggcgctttagtgcactctttaataatcaagcagttg 788
Db 2075 CCTGTATCAGTGGGCTTTGGGGCCACCATTATGGCTACGGCAACGAAGTCTACT 2016
QY 789 cgttggcgctgcaggttcttgggtgttcttattgatgtccagatatggtcatgt 848
Db 2015 CCTTTGGTTTGGCAAGTATTTTACCTTCATGCAAAACCATCCCGTCAACGGGAATTGATT 1956
QY 849 tcttggtgtcagttgttacctcttcctcagcattcgccagcagcagcattggtgccc 908
Db 1955 TCACCGCTCTGGGCCAGCGTATTTGGGGGTGTCATTGCCATCGGTTCGCGATTGTTCGGTA 1896
QY 909 ttacttggttcgcgcgaacgcagcagcattgatccagatgcaaacccgtgctccagtgccg 968
Db 1895 CGGTGATGCTTCATTTTCATCACCGTAAACGTCAGCCACGGCAGGTCGCC----- 1842
QY 969 caggaaacacccagcgaagcagagcagccgcagaaattttcaaacgattccaccatca 1028
Db 1843 -----GAAGAGAAACACAGAGGTTTATACACCACCTGAGCAGGGCGGTA 1798
QY 1029 tccaggcaccttggcgggtgaagctattgcactgagcagcgtcagcagtcagcagttgtg 1088
Db 1797 TGTGTTTACCGGATGAGGGAGAGATGTCGCTCATTCACGTCGCTGATACCACTGTTG 1738
QY 1089 ccagcggaaagcttgctcgggcttgccatcgctcccaacaaaggcagttagtttctc 1148
Db 1737 CCAGTGGCCTGTGGTAAAGTATTTGCCATTCCTGCGTGGTGAAGTCGGTTC 1678
QY 1149 cggtagtggaagatttgggtggcaatcccatctggccatgctttcgcagttcgcacca 1208
Db 1677 CGGTTGCGGGTCGAATTTGCTTTCGCCACATFACA-----CGCCATTGGCATG 1624
QY 1209 aggtcagagatggttccaatgttgatatattgatgcacattggtttcgacagtaaac 1268
Db 1623 AGTCAGATGATG-----TGTGGAGATCCTGATTATCGGTATCGACACCGGTAAAC 1570
QY 1269 tcaacgcgcagcacttaaccgcgtgaagaagcagcagcagatgaagtaaaagcagggagc 1328
Db 1569 TGGACGGCAAAATCTTTTCGGCTCAGCTCAGCTGGGTGACAAAGTCAATACAGGGCATC 1510
QY 1329 tgcgtgtgaattcgatattgatgccattaaaggctgcaggttatgaggttaaccacgcga 1388
Db 1509 GGCTGATTTCTTTTGATATCCCTGCTATTTCGGCAGGCGGATTTGATCTGACGACGCGG 1450
QY 1389 ttgtgtttcgaaattacaagaaacgcgcctgttaaacacttacggtttggcgcaaatg 1448
Db 1449 TATTAATCAGTAATAGCGATGATTTTACGGACGATATTACCCACGGCAGCGCGAGATAA 1390
QY 1449 aagcgggagccaaacctgctcaacgtc 1474
Db 1389 GGCAGGTGAACCGCTGTTATCCATC 1364

```

RESULT 5

```

US-08-920-828-17/c
; Sequence 17, Application US/08920828
; Patent No. 585398
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien

```

```

; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448

```

```

; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA

```

```

; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24

```

```

; US-08-920-828-17

```

Query Match 7.2%; Score 109.6; DB 2; Length 3615;

Best Local Similarity 46.3%; Pred. No. 2.2e-23; Mismatches 589; Indels 37; Gaps 4;

Matches 540; Conservative 0;

QY 312 tcaactcagtgctgacgttgctgctcaccgggattcttaccattcaccgcatggccag 371

Db 2495 TCACACCATTTGCTATGCTGATGGTTATCACACCGCTCAGCTTTCCTGCTGGGGCGC 2436

QY 372 caatgcgctggggtggcgatgtgctggcagcaggtctacagggacatttatgattcggtg 431

Db 2435 TATCAACCTGGTAAGCAACTGATTGCCGCGGTTATCTCTGGCTTTATCAGCGGTTT 2376

QY 432 gtccagctggcggtctgctctcggtctgcttactcaccacacacacacacacacacac 491

Db 2375 CTGCATTTGGCGGCGCGGTAATGGCGGCTTCGGCAAAATCTTCGTCATCTTCGACTGC 2316

QY 492 accagctcttccgcacaaattgagctggagctgtttaaccagggtgagctcttcatctcg 551

Db 2315 ACTGGGGCCTGGTGGCGGTGTGTATCAATAACTTTCACCGTGTCTGGGCTACGACACCATGA 2256

QY 552 caacggcatctatg---gctaataatgcggcaggggtggcgcatgtttggcagttgttcttc 608

Db 2255 TCCCGCTGTTAATGCCCGCATTTATGGCGCAGGTTCGGGGCGGCTCGGCGCTCTCTCTCT 2196

QY 609 tggcgaagagtgaaaaagctcaaggccttgcaggtgttcaggtgttcaggtgtctcgcgtgttc 668

Db 2195 GCGAACGCGATGCGCAAGAAAAGTGGTGGCGGGATCAGCGGCGTTCAGCAGGTCTGTTTG 2136

QY 669 gtattacgagcctgcgactcttcggtgtgaaccttcgctgcgctggcgcttcttcttcacg 728

Db 2135 GTATCACCAGAACCGGATATATGGCGTCAACCTGCCGCGTAAGTACCCCTTTGTTATCG 2076

QY 729 gtatcggtaccgcagctatcggtggcgctttgattgacctcttataatacaagcaggtg 788

Db 2075 CCTGTATCAGTGGGCTTTGGGGCGCACCATTTATGGCTACGGCAACCAAGAGTCTACT 2016

QY 789 cgttggcgctgcaggtttcttgggtgtgtttctattgatgtccagatatgtatgtatgt 848

Db 2015 CCTTTGGTTTGGCAAGTATTTTACCTTCATGCAAAACCAATCCCGTCAACGGGAATTGATT 1956

QY 849 tcttggtgtgcagttgtttaccttcttcttcacgtcattcggcgagcagcagcagcagcagc 908

Db 1955 TCACCGTCTGGCCAGCGTTATTGGCGGTGTCTATTGCATCGGTGGCGATTTCGGGTA 1896
QY 909 ttacttggttcgcgcgaacgagcattgatccagatgaacgcgtctccagtgctg 968
Db 1895 CGGTGATGCTTCATTTATCACCCTTAACGTCAGCCAGCGAGGTGCCCC----- 1842
QY 969 caggaacgacccaagcgaagcgaagcaccgcgcgcgaattttcaaacgattccaccatca 1028
Db 1843 -----GAAGAGAAACACACAGAGGTTATTACACCACCTGAGCAGGGCGGTA 1798
QY 1029 tccaggcacctttgacgggtgaagctattgcactgagcagcgtcagcgatgccaattgtt 1088
Db 1797 TCTGTTCCAGCATGACGGGAGAGATTGTGCTCTCATTCAGCTGCTGATACACAGTTTG 1738
QY 1089 ccagcgaagctggtcggggttcaccatgctcccaacgaagggcagttgattctc 1148
Db 1737 CCAGTGCCCTGTGGGTAAGGATTCCTCATTCGCCCTCGGTGGTGAAGTGGCTCTC 1678
QY 1149 cgtgagtgaaagattggtggaattcccatctgcccattggttcgagttcgacca 1208
Db 1677 CGGTGCGGGTCGAATTGCTTCGTTGTTCCGCACATTACA-----CGCCATTGGCATTG 1624
QY 1209 aggtgaggtggttccaaatggtgatactttgatgcacattggttttcgacacagtaaac 1268
Db 1623 AGTCAGATGATG-----TGTGGAGATCTCTGATTCATGTCGTFATCGACACCGTAAAC 1570
QY 1269 tcaacgacgcacactttaaccgctgaagcagcggcgatgaagtcaagcaggggagc 1328
Db 1569 TGGACGGCAAAATCTTTTCGCTCACCTCAACGTGGGTGACAGGTCATACAGCGCATC 1510
QY 1329 tctgtgtgaattgatattgatgcttaagcgtcaggttatgaggttaaccacgcgga 1388
Db 1509 GGCTGATTTCTTTTGAATCCCTGCTATTTCGAGGCGCGGATTGATCTGACGACGCGG 1450
QY 1389 ttgtgttggaattacaagaacacgacgttaaacacttaacgtttgtggcgaaattg 1448
Db 1449 TATTAATCAGTAATAGCGATGATTTTACGACGATTAATACCCACGCGCAGCAGATAA 1390
QY 1449 aagcgggagccaactctcaacgtc 1474
Db 1389 GCGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 6
US-08-673-190A-3/c
: Sequence 3, Application US/08673190A
: Patent No. 5985668
: GENERAL INFORMATION:
: APPLICANT: Mattes, Ralf
: APPLICANT: Klein, Kathrin
: APPLICANT: Stegmaier, Sabine
: TITLE OF INVENTION: Sucrose Metabolism Mutants.
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/673,190A
: FILING DATE: 27-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Forman, David S.
: REGISTRATION NUMBER: 33,694
: REFERENCE/DOCKET NUMBER: 06473.0001-00000

TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)408-4000
: TELEFAX: (202)408-4400
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 465 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Protaminobacter rubrum
: US-08-673-190A-3

Query Match 6.4%; Score 97.8; DB 2; Length 465;
Best Local Similarity 53.5%; Pred. No. 2.7e-20;
Matches 204; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 145 cgcaccatgctgcggcgcaaatgccaatggtccctgtttggttttagatgttgcaca 204
Db 441 CGCTGGGGGCTCGCGGGGTTTCCACACCATGAATTTCTCGGACTCGATATCGGCAT 382
QY 205 agcgggttacagggaccggtgttctctgctgctggtggtgttcttggattcttgcaacgat 264
Db 381 GATCGGCTATCAGGGTACGGTGTTCGGGTGCTGCTGACGGGTGTTTATGAGCGTGT 322
QY 265 cgagaagtcttcacaaagcagctcaaggcagctgcagacttctgatcactccagtgct 324
Db 321 GGAACAAACAGCTCGCGAGAGTATCCCAACGCGCTGACCTGATCTGACGCTTCTCT 262
QY 325 gacgttctgctcacggattctctacatcatcgcctggtggtggtggtggtggtggtggtggt 384
Db 261 GACGGTGGTATCTCCGGCTTCGTCGCCATGCTGTTTATCGGCCCGCGCGGCGCACACT 202
QY 385 gggcgatgctgctggcacacggtctacagggaactttatgatttgcgtggtggtccagtcggcg 444
Db 201 GGGTGACGGCATCTCCTTCGTCGTCAGCAGCTGATTGCCACGCCGCTGTTGTTTCGCGG 142
QY 445 tctgctcttcgggtctgctactcaacgaatgcacagctgcacgtggtggtggtggtggtggtggt 504
Db 141 ATTCTGTTCGGCGGCTGTATTCCGCCATCGTCAATCAGCGCATCACCAGCATTCCCA 82
QY 505 gccaatgagctgagctgtt 525
Db 81 CGCGTAGAGGGGGGCTGCT 61

RESULT 7
US-08-673-190A-6
: Sequence 6, Application US/08673190A
: Patent No. 5985668
: GENERAL INFORMATION:
: APPLICANT: Mattes, Ralf
: APPLICANT: Klein, Kathrin
: APPLICANT: Stegmaier, Sabine
: TITLE OF INVENTION: Sucrose Metabolism Mutants.
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/673,190A
: FILING DATE: 27-JUN-1996


```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      2.4%; Score 37.2; DB 4; Length 4403765;
Best Local Similarity 43.7%; Pred. No. 18;
Matches 165; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 627 tcaaggccctgcaggtgctccaggtgtctccgctgttcttggtattacagagccctgcga 686
Db 3923279 TTACCGCGCTGCGCGCGCTTGGCCCGCTGCGCGCGATGCCCTGGCTGCCAGCGTTACCG 3923320

QY 687 tcttcggtggaaccttcgctgcgtgcgttcttcctcatcggtatcggtaccgcagcta 746
Db 3923219 CCGGTAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCTTCCGCTGCCA 39233160

QY 747 tcggtggcgctttgattgcaactctttaataatcaaggcagttgctggtggcgctgcaggtt 806
Db 3923159 CCGTTCGCCACCGCGCGCGCGCTTGGGTGGGCAAGCTGTGGCGCGGAGCACTGGCGG 39233100

QY 807 tcttggtgtgtttctattgatgtctccagatggtcatgttcttctgtgtgctgtg 866
Db 3923099 CCGACGCTGCGGTGGCGCGCGCTTACCGCGCTCACCGCGCTTGGCGCGCTTGGCGCG 39233040

QY 867 ttaccttctcatcgcaattcggcgcaagcagattgcttatgaccttacttctgttgcgca 926
Db 3923039 CCACCGTTCGCCCGCGCTAAGCGCGGTGGCGCGGTTCGCCCGCGCACCGCGCTACCG 3922980

QY 927 acggcagcattgatccagatgcaaccgctgctccagtgctgcaggaacaccaaagccg 986
Db 3922979 GTCCCGCGCTTGGCGCGCTTGGCGCGCTTGGCGGAGCGCGCGCTGGTGTATTGCCACCG 3922920

QY 987 aagcagaagcaccgcag 1004
Db 3922919 AAGTCGCGGAGCGCGCG 3922902

RESULT 10
US-08-471-119A-1/c
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoerendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-1

Query Match      2.3%; Score 34.6; DB 1; Length 46899;
Best Local Similarity 55.4%; Pred. No. 9.4;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 61 cgttcggcggaatgagttcctggcgccgcgcgtattggtatggcgtatggttcccgag 120
Db 29376 CGATACGGCGGCTCTGAGTCGTGGAGTGGCGGTTTCGTACGAGTTTCGAAGTTTCGAC 29317

QY 121 ctgttgtaacggtacgactgacgtgcccgcaccatggtgcggcggaatgcaatgtggtc 180
Db 29316 CTGTGATGGTGGTAGGAGAGTGTACAGGACACGACTGCCCTGCCAATTTCCATCGGTTG 29257

QY 181 c 181
Db 29256 C 29256

RESULT 11
US-08-460-751-1/c
; Sequence 1, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Glucksmann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 22, 2002, 06:40:16 ; Search time 1607.29 Seconds
(without alignments)
10208.988 Million cell updates/sec

Title: US-09-604-231-1
Perfect score: 1527
Sequence: 1 ctcacggcatctgcgcgtt.....gttgaacacctgagtgttcg 1527

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estcom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: gb_estl: *
11: gb_est2: *
12: gb_htc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pin: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
c 1	48.8	3.2	680	10	AL506262
c 2	45.8	3.0	473	11	T03017
c 3	41.8	2.7	681	13	CNS02EOD
4	41.6	2.7	895	13	CNS0071A
5	41.4	2.7	477	11	BE997958
6	41.4	2.7	559	10	AW574083
7	41.4	2.7	601	11	BC580528
8	41.4	2.7	627	11	BE998413
9	40.8	2.7	884	13	CNS00600
c 10	40.8	2.7	983	13	CNS03CVD
11	40.8	2.7	1101	13	CNS00LXJ
c 12	40.2	2.6	970	13	CNS03H6V

13	39.8	2.6	606	11	BE997957
c 14	39.2	2.6	993	13	CNS00LOP
c 15	38.8	2.5	539	10	AI070146
16	38.8	2.5	714	10	AW583970
c 17	38.6	2.5	401	10	AI436816
c 18	38.6	2.5	1204	11	BC391833
19	38.4	2.5	1101	13	CNS0021D
c 20	38.2	2.5	500	11	BF727921
c 21	38.2	2.5	525	13	CNS025EN
22	38.2	2.5	531	10	AW573719
23	38.2	2.5	621	11	BE998387
c 24	38.2	2.5	652	10	AI982977
25	38.2	2.5	687	10	AW980716
26	38.2	2.5	703	10	AW980735
27	38.2	2.5	713	11	BC581619
28	38.2	2.5	737	11	BC581892
29	38.2	2.5	789	11	BC582423
30	38.2	2.5	795	11	BC583630
c 31	37.8	2.5	654	10	AI812147
c 32	37.6	2.5	780	11	BI118076
33	37.6	2.5	1101	13	CNS00FXE
c 34	37.4	2.4	460	10	AI318021
35	37.4	2.4	552	10	BE721200
c 36	37.4	2.4	904	13	CNS03HA0
37	37.2	2.4	421	10	AW173566
38	37.2	2.4	561	10	AI917973
39	37.2	2.4	594	10	AW440291
40	37.2	2.4	622	10	AI499080
c 41	37.2	2.4	647	10	BE158004
42	37.2	2.4	650	10	AW516253
43	37.2	2.4	710	10	AI564728
44	37.2	2.4	906	13	CNS02GJN
45	37	2.4	440	13	AZ982613

ALIGNMENTS

RESULT 1
AL506262/c
LOCUS AL506262 680 bp mRNA EST 04-JAN-2001
DEFINITION AL506262 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
Hordeum vulgare cDNA clone HY02124T 5', mRNA sequence.
ACCESSION AL506262
VERSION AL506262.1 GI:12032477
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 680)
AUTHORS Michalek W., Weschke W., Pleissner K.-P. and Graner A.
TITLE EST sequencing and analysis in barley
JOURNAL Unpublished (2000)
COMMENT Contact: Michalek W
Institute for Plant Genetics and Crop Plant Research
Corrensstr.3, D-06466 Gatersleben, Germany
Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
Seq primer: T3 primer for 5'end.

FEATURES

Location/Qualifiers
1..680
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY02124T"
/clone_lib="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"
/tissue_type="developing caryopsis (3.-15.DAP)"
/lab_host="XLOR"
/note="Vector: plasmid pBK-CMV; Site.1: EcoRI; Site.2: XhoI; mRNA was made from developing caryopsis (3.-15.DAP) of spring barley variety 'Barke', a high quality malting

REFERENCE 3 (bases 1 to 681)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source Location/Qualifiers
 1. .881
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="262H14"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0A262DD07LP1-end : T7"

BASE COUNT 74 a 105 c 170 g 282 t 50 others
 ORIGIN

Query Match 2.7%; Score 41.8; DB 13; Length 681;
 Best Local Similarity 40.9%; Pred. NO. 2.2;
 Matches 139; Conservative 18; Mismatches 183; Indels 0; Gaps 0;

QY 632 ggccgtgcagggtcctcaggtctccgctgtcttcttgattacgagcctgcgacatcttc 691
 Db 58 GCGGTGTGTTGTTTGTGTTGTTGTCGCGYGTGTTGTTGTTTGTGTTGTT 117
 QY 692 ggtgtgaacctgcctgcgcgtgcccgtttcttcattcaggtatcggtaccgcagcatcgg 751
 Db 118 GTTGTGTTTGTGTTGTTGTTGTTGTTGTCGCGYGTGTTGTTGTTTGTGTT 177
 QY 752 ggcgcttgatgcactctttaaatacaagcagtgctggtggcgctcaggtttcttg 811
 Db 178 GTNGTGTGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 237
 QY 812 ggtgtgtttctattgatctccagataggtcatgtcttcttggtgtgcagttgttacc 871
 Db 238 GTGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 297
 QY 872 ttcttcacgcatcggcgcagcagtgattgcttaccgttacttgctgcgcgaacggc 931
 Db 298 GCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 357
 QY 932 agcattgaccagatcaaacccgtcgtccagtgccagtcag 971
 Db 358 GTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 397

RESULT 4
 CNS0071A 895 bp DNA GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL066286
 VERSION AL066286.1 GI:4945153
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 895)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .895
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR14B09"
 /note="end : TET3"

BASE COUNT 124 a 80 c 204 g 179 t 308 others
 ORIGIN

Query Match 2.7%; Score 41.6; DB 13; Length 895;
 Best Local Similarity 19.9%; Pred. NO. 2.7;
 Matches 68; Conservative 115; Mismatches 159; Indels 0; Gaps 0;

QY 580 ggggtgcggcatgtttggcagtggtttctctgcgcaagagtgaaagctcaaggcccttgc 639
 Db 512 GKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGK 571
 QY 640 aggtgcttcaggtgctcgcgtgttcttgattacgagcctgcgcatcttcggtgtgaa 699
 Db 572 KKTGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 631
 QY 700 ccttcgcctgcctggccgttcttcacgttatcggtatcggtacgcagctatcggtgccttt 759
 Db 632 KGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 691
 QY 760 gattgcactctttaaatacaagcagtgctggtggcgctgcaggtttcttgggtttgt 819
 Db 692 KKTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 751
 QY 820 ttctattgatgccagatagtgatcttctgtgtgctgagttgttaccttctcat 879
 Db 752 KGKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 811
 QY 880 cgcattcggcgagcagtgattgcttatggcctttacttctggttcg 921
 Db 812 TTTGTGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 853

RESULT 5
 BE997958 477 bp mRNA EST 06-OCT-2000
 LOCUS EST429681 GVSU Medicago truncatula cDNA clone pcVSN-8B9, mRNA
 DEFINITION sequence.
 ACCESSION BE997958
 VERSION BE997958.1 GI:10698234
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
 1 (bases 1 to 477)
 Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town,C.D., Bowman,C.L., Craven,M.B., Cho,J., Fraser,C.M.
 ESTs from senescent nodules of Medicago truncatula
 Unpublished (2000)
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA

Tel: 612 625 5715
 Fax: 651-649-5058
 Email: vance004@maroon.tc.umn.edu
 University of Minnesota name: M271936e TIGR sequence name:
 MTKAK05TKB More information is available at:
 http://chrysie.tamu.edu/medicago
 Seq primer: SKmod (CTA gAA CTA gtg gat CC).

FEATURES

source

1. .477
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pGVSN-8B9"
 /clone_lib="GVSN"
 /tissue_type="senescent root nodules"
 /dev_stage="mixture of effective nodules from 40 day old
 plants harvested 36 hours post shoot removal and nodules
 collected from 2 month old plants at mid-pod stage"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from the
 mixture of effective nodules of 40 day old plants
 harvested 36 hours post shoot removal and nodules
 collected from 2 month old plants at mid-pod stage. The
 cDNA was directionally ligated into the Uni-ZAP XR vector
 from Stratagene and packaged using Gigapack III Gold
 packaging extracts. Plasmids containing cDNA inserts were
 excised from the recombinant lambda-ZAP phage using
 Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT

ORIGIN

137 a 100 c 101 g 139 t
 Query Match 2.7%; Score 41.4; DB 11; Length 477;
 Best Local Similarity 50.2%; Pred. NO. 2.6;
 Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1246 cattggtttccgacagtaaacctcaacgcgcgcactttaaccgcgtgaagcagcggg 1305
 DB 258 CATGACTTCACCAAGAAGAGCTTCTCATAGCTTTCTTAAGCCTAACAGGATCAATAA 317
 QY 1306 cgatgaagtcacaaagcagggagctgctgtgtaattcgattgatgccattagcgtgc 1365
 DB 318 TGGTTTTGTCATCGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAGCGGAC 377
 QY 1366 aggttatgaggttaaccacgcgcgattgtgttcgaattacaaagaaacccgacctgtaaa 1425
 DB 378 CAATGAAGAAGTGAGCGACTTTGTTCTGAGGCTGATTTCTGATAGTATGGACATCTAAG 437
 QY 1426 cacttacggtttggcgaaattg 1448
 DB 438 CTTTAAGGAGTTTGTGACAGACTTG 460

RESULT 6

AW574083

LOCUS

EST316674 GVN Medicago truncatula cDNA clone pGVN-51G2, mRNA

DEFINITION

sequence.

ACCESSION

AW574083.1 GI:7238816

VERSION

EST.

KEYWORDS

SOURCE

barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

REFERENCE

1 (bases 1 to 559)

AUTHORS

Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, J.S., Peng

H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,

Holt, I.E. and Fraser, C.M.

ESTs from one month old nitrogen-fixing root nodules of Medicago

JOURNAL

COMMENT

truncatula
 Unpublished (2000)
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: vance004@maroon.tc.umn.edu
 Minnesota EST name: M252981e
 TIGR sequence name: MTCAR37TK
 More information is available at:
 'http://chrysie.tamu.edu/medicago'
 Seq primer: SKmod (CTA gAA CTA gtg gat CC).

FEATURES

source

1. .559
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pGVN-51G2"
 /clone_lib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month
 post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 effective root nodules harvested one month post
 inoculation with Sinorhizobium meliloti. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XL0LR cells."

BASE COUNT

ORIGIN

159 a 112 c 112 g 176 t
 Query Match 2.7%; Score 41.4; DB 10; Length 559;
 Best Local Similarity 50.2%; Pred. NO. 2.7;
 Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1246 cattggtttccgacagtaaacctcaacgcgcgcactttaaccgcgtgaagcagcggg 1305
 DB 316 CATGACTTCACCAAGAAGAGCTTCTCATAGCTTTCTTAAGCCTAACAGGATCAATAA 375
 QY 1306 cgatgaagtcacaaagcagggagctgctgtgtaattcgattgatgccattagcgtgc 1365
 DB 376 TGGTTTTGTCATCGCATCTGAGTTGCATTATTACTTGACAAATCAAGGCATTAAGCGGAC 435
 QY 1366 aggttatgaggttaaccacgcgcgattgtgttcgaattacaaagaaacccgacctgtaaa 1425
 DB 436 CAATGAAGAAGTGAGCGACTTTGTTCTGAGGCTGATTTCTGATAGTATGGACATCTAAG 495
 QY 1426 cacttacggtttggcgaaattg 1448
 DB 496 CTTTAAGGAGTTTGTGACAGACTTG 518

RESULT 7

BG580528

LOCUS

EST482254 GVN Medicago truncatula cDNA clone pGVN-57H8 5' end, mRNA

DEFINITION

sequence.

ACCESSION

BG580528.1 GI:13595592

VERSION

EST.

KEYWORDS

SOURCE

barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

REFERENCE 1 (bases 1 to 601)
 AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town
 C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
 TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago
 truncatula, 2001
 JOURNAL Unpublished (2001)
 COMMENT Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: vance004@maroon.tc.umn.edu
 University of Minnesota name: M381580e TIGR sequence name:
 MTCT40TK More information is available at: http://www.medicago.org
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

FEATURES
 Location/Qualifiers
 1..601
 /organism="Medicago truncatula"
 /cultivar="genotype Al7"
 /db_xref="taxon:3880"
 /clone="pGVN-57H8"
 /clone_lib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month
 post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 effective root nodules harvested one month post
 inoculation with Sinorhizobium meliloti. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XL0LR cells."
 BASE COUNT 170 a 121 c 127 g 183 t
 ORIGIN

Query Match 2.7%; Score 41.4; DB 11; Length 601;
 Best Local Similarity 50.2%; Pred. No. 2.7;
 Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1246 cattgtttcgcacagtaaacctcaacgagcagcacttaaccgctgaagaagcagg 1305

Db 307 CATTGACTTCACCAAGAAGAGAGCTTCTCATAGCTTTCTTAAGCCCTAACACGGATCATAA 366

QY 1306 cgatgaagtcacagcagggagctgtgtgaattcgatattgatgccattagcctgc 1365

Db 367 TGGTTTGTGCTGCTGATCTGAGTTGCTGATTTACTTGACAAATCAAGGATTAAGCGGAC 426

QY 1366 aggttatagggttaaccacccgctgtgttttcgaattacaagaacccgacctgtataa 1425

Db 427 CAATGAAGAAGTGAGGAGCTTTTGTCTGAGGCTGATTTCTGATGATGATGATCTAAG 486

QY 1426 cacttacggtttggcgaaattg 1448

Db 487 CTTTAAGGAGTTTGTGACAGACTTG 509

RESULT 8
 BE998413 627 bp mRNA EST 06-OCT-2000
 LOCUS EST430136 GVSN Medicago truncatula cDNA clone pGVSN-9J12, mRNA
 DEFINITION sequence.

ACCESSION BE998413
 VERSION BE998413
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 627)
 AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
 C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
 TITLE ESTs from senescent nodules of Medicago truncatula
 JOURNAL Unpublished (2000)
 COMMENT Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: vance004@maroon.tc.umn.edu
 University of Minnesota name: M272391e TIGR sequence name:
 MTKAP54TK More information is available at:
 http://chrysis.tamu.edu/medicago
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

FEATURES
 Location/Qualifiers
 1..627
 /organism="Medicago truncatula"
 /cultivar="genotype Al7"
 /db_xref="taxon:3880"
 /clone="pGVSN-9J12"
 /clone_lib="GVSN"
 /tissue_type="senescent root nodules"
 /dev_stage="mixture of effective nodules from 40 day old
 plants harvested 36 hours post shoot removal and nodules
 collected from 2 month old plants at mid-pod stage"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from the
 mixture of effective nodules of 40 day old plants
 harvested 36 hours post shoot removal and nodules
 collected from 2 month old plants at mid-pod stage. The
 cDNA was directionally ligated into the Uni-ZAP XR vector
 from Stratagene and packaged using Gigapack III Gold
 packaging extracts. Plasmids containing cDNA inserts were
 excised from the recombinant lambda-ZAP phage using
 Ex-Assist helper phage and propagated in SOLR cells."
 BASE COUNT 175 a 124 c 135 g 193 t
 ORIGIN

Query Match 2.7%; Score 41.4; DB 11; Length 627;
 Best Local Similarity 50.2%; Pred. No. 2.8;
 Matches 102; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1246 cattgtttcgcacagtaaacctcaacgagcagcacttaaccgctgaagaagcagg 1305

Db 313 CATTGACTTCACCAAGAAGAGAGCTTCTCATAGCTTTCTTAAGCCCTAACACGGATCATAA 372

QY 1306 cgatgaagtcacagcagggagctgtgtgaattcgatattgatgccattagcctgc 1365

Db 373 TGGTTTGTGCTGCTGATCTGAGTTGCTGATTTACTTGACAAATCAAGGATTAAGCGGAC 432

QY 1366 aggttatagggttaaccacccgctgtgttttcgaattacaagaacccgacctgtataa 1425

Db 433 CAATGAAGAAGTGAGGAGCTTTTGTCTGAGGCTGATTTCTGATGATGATGATCTAAG 492

QY 1426 cacttacggtttggcgaaattg 1448

Db 493 CTTTAAGGAGTTTGTGACAGACTTG 515

RESULT 9
 CNS006U0 884 bp DNA GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
 DEFINITION BAC14N21 of RPC1-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL065923
 VERSION AL065923.1 GI:494891

```

KEYWORDS  GSS.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster

REFERENCE  1 (bases 1 to 884)
AUTHORS    Direct Submission
TITLE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL    BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT    - Web : www.genoscope.cns.fr
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mamozer in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2: cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES   Location/Qualifiers
            source
            1..884
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"
            /clone="BACR1421"
            /note="end : 77"

BASE COUNT 230 a 62 c 139 g 124 t 329 others
ORIGIN

Query Match      2.7%; Score 40.8; DB 13; Length 884;
Best Local Similarity 15.0%; Pred. No. 4.4;
Matches 55; Conservative 147; Mismatches 165; Indels 0; Gaps 0;

QY 572 atgcgccagggtgcagatgttggcagtggtttcttcctgcggaagagtgaaagctcaag 631
Db 516 RKGVGMVAVKTTGGGGGKRTGKKGSKSTKGDGKGVGTGKTKTTKDDTTGKKTGG 575

QY 632 ggccttcaggtgcttcaggtgtctcogctgttcttggattacggagcctgcgatcttc 691
Db 576 TGAGKTGGKGGKGGTGGGTGGGKGGKGGTGTGTTKTKTKTKTKTKTKTKTKTKTKTK 635

QY 692 ggtgtgaaccttcgcctgcgcgtcttcttcacgttcggtaccgcagctatcggt 751
Db 636 GGGGGGGKKGKGGTGCGGGGKKGKGTGTRGDRDCTGCKKTKTKKKKKKKGKKGK 695

QY 752 ggcgcttgattgcaactcttaataatacaaggcagttgcgttggcgctgcaggtttctt 811
Db 696 KGTGKTKKSGTKKKGKGTGKGTGKGTGKTKKKGKGGKGGKKGKKGKKGKKGK 755

QY 812 ggtgtgttcttattgatgtccagatattggtcatgttcttcttgggtgtgagttgtacc 871
Db 756 KKKKGGKKKTKTKGKGGTKKGGTKKGGTKKGGTKKGGTKKGGTKKGGTKKGGTKKGG 815

QY 872 ttcttcacgcatcgcgcagcagattgcttattggtcttacttggcttcgcccgaacg 931
Db 816 KKKKKKKTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 875

QY 932 agcattg 938
Db 876 KKBKKBK 882

RESULT 10
CNS03CVD/c

```

```

LOCUS      CNS03CVD      983 bp      DNA      GSS      15-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            015D07 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL238306
VERSION    AL238306.1 GI:7897441
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 983)
AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bounau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
            Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
            Unpublished
            2 (bases 1 to 983)
            Roest-Crollius,H., Jaillon,O., Dasilva,C., Bounau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
            Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
            Unpublished
            3 (bases 1 to 983)
            Genoscope.
            Direct Submission
            Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.

FEATURES   Location/Qualifiers
            1..983
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="015D07"
            /clone_lib="G"
            /note="Genoscope sequence ID : COBG015CB04SP1-end :
            PUC-Ori"

BASE COUNT 202 a 289 c 274 g 201 t 17 others
ORIGIN

Query Match      2.7%; Score 40.8; DB 13; Length 983;
Best Local Similarity 53.0%; Pred. No. 4.5;
Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 705 gctgcgctggcgttcttcacgtatcggtaccgcagctatcggtggcgcttgattg 764
Db 732 GCCGAGCCGGGAGCGGCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 673

QY 765 cactcttaataatacaaggcagttgcgttggcgctgcaggttcttctgggtgttctcta 824
Db 672 TGTATTATTATCATATTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 613

QY 825 ttgatctccagatattgctcatgttcttcttgggtgtgctgagttgtt 868
Db 612 TTGTTGCGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 569

RESULT 11
CNS00LXJ
LOCUS      CNS00LXJ      1101 bp      DNA      GSS      14-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR48E16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL078875
VERSION    AL078875.1 GI:5102165
KEYWORDS   GSS.

```

[illegible]

VERSION		AL243904.1 GI:7964916					
KEYWORDS		GSS; genome survey sequence.					
SOURCE		Tetraodon nigroviridis.					
ORGANISM		Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. 1 (bases 1 to 970)					
REFERENCE		Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis Unpublished					
JOURNAL							
REFERENCE		2 (bases 1 to 970) Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence Unpublished					
TITLE		3 (bases 1 to 970) Genoscope.					
JOURNAL							
REFERENCE		Direct Submission Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .					
FEATURES		Location/Qualifiers 1..970 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone_xref="026011" /clone_lib="G" /note="Genoscope sequence ID : COBG026AH06LP1-end : T7"					
BASE COUNT		311 a 298 c 194 g 152 t 15 others					
ORIGIN							
		Query Match 2.6%; Score 40.2; DB 13; Length 970;					
		Best Local Similarity 45.9%; Pred. No. 6.5;					
		Matches 130; Conservative 2; Mismatches 151; Indels 0; Gaps 0;					
QY	637	tgcaggtcttcagggtgtctccgcgtgttcttggtattacaggacctcgcgatcttcggtgt	696				
Db	709	TGCTGCTGTGTTCCTCCTCCTCTGTGTGTGTGTGTGCCTGCTGCTGTGTGTGCTGT	650				
QY	697	gaaccttcgcctgcctgcccgtcttcctcatcggtatcggtaccgcagctatcggtggcgc	756				
Db	649	TGTTGCTGCTGCTGCTGCTGTTGTTGTTGTVGVGCTGCTGTTGTTGAATGACTGCTGTTGTTGT	590				
QY	757	tttgattgcactctttaatatcaaggcaggttgcgttggcgcgctgcaggtttcttggsgtgt	816				
Db	589	TGCTGCTGCTGCTGTTGTTGTTGMCTGCTGTTGTTGTTGCTGCTGCTGCTGTTGTTGCTGC	530				
QY	817	tgttcttatgatgctccagatatggctcatgttcttggtgtgtgcagtgttaacctctt	876				
Db	529	TGCTGTTGTTGTTGCTGCTGCTGTTGTTGCTGCTGTTGTTGCTGCTGTTGTTGCTGTTGT	470				
QY	877	cattcgattcggcgcagcagttgcttatggccttacttggtt	919				
Db	469	TGTTGCTGCTGCTGCTGCTGTTGTTGCTGCTGCTGTTGTTGCTGCTGTTGTT	427				
RESULT 13							
BE997957							
LOCUS		BE997957 506 bp mRNA EST 06-OCT-2000					
DEFINITION		EST429680 GVSN Medicago truncatula cdna clone pgVSN-8B9, mRNA sequence.					
ACCESSION		BE997957					
VERSION		BE997957.1 GI:10698233					

KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 606)
AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P., Town C.D., Bowman, C.L., Craven, M.B., Cho, J., and Fraser, C.M.
TITLE ESTs from senescent nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M271935e TIGR sequence name: MTKAK05NK
More information is available at: http://chrysie.tamu.edu/medicago
Seq primer: Sknod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers
1..606
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVSN-889"
/clone_lib="GVSN"
/tissue_type="senescent root nodules"
/dev_stage="mixture of effective nodules from 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"
/lab_host="E. coli strain SOLR"
/note="vector: pBluescript SK +/-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

BASE COUNT 173 a 122 c 127 g 184 t
ORIGIN

Query Match 2.6%; Score 39.8; DB 11; Length 606;
Best Local Similarity 49.8%; Pred. No. 7.3;
Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1246 cattggtttcgacacagtaaacctcaacgcgcacactttaaccctgaagacgagg 1305
Db CATTTGACTTACCACAAAGAGAGCTTCTCATAGCTTCTCTAAGCCTTAACACGGATATAA 376
QY 1306 cgaatgaagcaaacgaggagctgtgtgaattcgatattgatgccattagctgc 1365
Db TGGTTTGTGCATGTCATGTGAGTTGATATTACTTGCACAAATCAAGGCAATTAAGCGCAC 436
QY 1366 aggttatgaggttaacacgcgcgattgttttcgaattacaagaaaaccgacacctgtaa 1425
Db CAATGAAGAGTGAAGCGACTTTGTCGTAGGCTGATTCGTATGATGGACATCTAAG 496
QY 1426 cacttacggtttggcgcaattg 1448
Db CTTTAAAGAGCTTGTTCAGACTTG 519

RESULT 14
CNS00L0P/c 993 bp DNA GSS 03-JUN-1999
LOCUS

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR23F02 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL067821
VERSION AL067821.1 GI:4957842
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 993)
REFERENCE Direct Submission
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
TITLE BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by Partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
1..993
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR23F02"
/note="end : T7"

BASE COUNT 379 a 207 c 186 g 151 t 70 others
ORIGIN

Query Match 2.6%; Score 39.2; DB 13; Length 993;
Best Local Similarity 48.4%; Pred. No. 12;
Matches 104; Conservative 1; Mismatches 110; Indels 0; Gaps 0;

QY 705 gcctgcgctggccgtttcttcacgttatcggtaccgcagctatcggtggcgcttgattg 764
Db 229 GCTGCGGTTTGGCGCTCCATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170
QY 765 cactctttaatatcaaggcagttgcgttggcgctgcaggtttcttgggtgtgtttctta 824
Db 169 CTGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 110
QY 825 ttgatctccagatatgggtcatgttcttgggtgtgtgcagttgttaacctcttcacatc 884
Db 109 TTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 50
QY 885 tggcgccagcagattgcttattggcctttacttggtt 919
Db 49 TTGTTGGCGCTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 15

RESULT 15
AI070146/c
LOCUS AI070146 539 bp mRNA EST
DEFINITION UI-R-Y0-lu-g-12-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
UI-R-Y0-lu-g-12-0-UI 3', mRNA sequence.
ACCESSION AI070146
VERSION AI070146.1 GI:3396397
KEYWORDS EST.
SOURCE Norway rat.

ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 539)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	9704477
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565

Email: mscores@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to identify it as a clone from the normalized adult Eye library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLNL (infoimage.llnl.gov). IMAGE ID=1785384 The following repetitive elements were found in this cDNA sequence: 357-476, >(CAA)n\$Simple_repeat 478-538, >(CAA)n\$Simple_repeat
Seq primer: M13 Forward

POLYA-No. Location/Qualifiers

1. .539

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-Y0-lu-g-12-0-UI"

/clone_lib="UI-R-Y0"

/dev_stage="adult"

/lab_host="PH10B (Life Technologies)"

/note="Vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-Y0 library is a subtracted library derived from an individually-tagged normalized whole-eye (minus the lens) library. The driver for the subtraction consisted of a pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

source

FEATURES

[illegible]

Qy 846 tgttcttggtgtagcagttgttaaccttcttcattcgccagcgattgcta 903
| | | | | | | | | | | | | | | | | |
Db 401 ttttttgttgatgtgcgtgtgtttaattctgtgtgtgtgtgtgtgtgtgt 344

Search completed: March 22, 2002, 08:11:36
Job time: 5480 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 09:23:24 ; Search time 66.97 Seconds
(without alignments)
3750.397 Million cell updates/sec

Title: US-09-604-231-3

Perfect score: 1109

Sequence: 1 tatgatttcggcggtccagt.....gttgaaccttgatttcg 1109

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	91	8.2	3615	1	US-08-920-812-17
C 2	91	8.2	3615	1	US-08-920-827-17
C 3	91	8.2	3615	1	US-08-921-177-17
C 4	91	8.2	3615	1	US-08-362-577C-17
C 5	91	8.2	3615	2	US-08-920-828-17
C 6	39.8	3.6	465	2	US-08-673-190A-3
C 7	39.6	3.6	7218	1	US-08-232-463-14
C 8	37.2	3.4	4403765	4	US-09-103-840A-2
C 9	32.6	2.9	477	4	US-09-135-994-1
C 10	32.4	2.9	292	4	US-09-117-121-29
C 11	32.2	2.9	7218	1	US-08-232-463-14
C 12	32	2.9	1008	4	US-09-198-955A-9
C 13	32	2.9	3533	3	US-08-985-916-15
C 14	32	2.9	5163	3	US-08-700-651-1
C 15	32	2.9	5163	3	US-08-928-361B-4
C 16	32	2.9	5318	3	US-08-700-651-2
C 17	32	2.9	5318	3	US-08-928-361B-3
C 18	32	2.9	10348	2	US-08-457-273B-41
C 19	32	2.9	10348	3	US-08-556-419-13
C 20	32	2.9	10348	4	US-09-041-886-14
C 21	32	2.9	10366	1	US-08-246-982A-5
C 22	32	2.9	10366	1	US-08-453-285-5
C 23	31.6	2.8	5511	3	US-08-928-361B-2
C 24	31.6	2.8	7334	3	US-08-928-361B-1
C 25	31.6	2.8	7655	1	US-08-619-554-1
C 26	31.2	2.8	342	4	US-08-905-223-196
C 27	30.8	2.8	2745	1	US-08-363-255-1

28	30.8	2.8	2745	1	US-08-363-255-13	Sequence 13, Appl
C 29	30.6	2.8	289	4	US-09-007-005-17	Sequence 17, Appl
C 30	30.6	2.8	289	4	US-09-244-796-17	Sequence 17, Appl
C 31	30.4	2.7	688	4	US-08-998-416-915	Sequence 915, App
C 32	30.4	2.7	3807	1	US-08-022-835-5	Sequence 5, Appl
C 33	30.4	2.7	3807	1	US-08-388-809-5	Sequence 5, Appl
C 34	30.4	2.7	3807	2	US-08-647-714-5	Sequence 5, Appl
C 35	30.4	2.7	20303	1	US-08-370-975B-6	Sequence 6, Appl
C 36	30.4	2.7	26764	1	US-08-370-975B-1	Sequence 1, Appl
C 37	30.4	2.7	19136	4	US-09-422-869-1	Sequence 1, Appl
C 38	30.2	2.7	1996	2	US-08-559-524A-1	Sequence 1, Appl
C 39	30.2	2.7	1996	3	US-08-749-707-1	Sequence 1, Appl
C 40	30.2	2.7	2277	1	US-08-676-967-2	Sequence 2, Appl
C 41	30.2	2.7	2277	1	US-08-676-974-2	Sequence 2, Appl
C 42	30.2	2.7	2277	2	US-09-098-487-2	Sequence 2, Appl
C 43	30.2	2.7	4692	2	US-08-916-917-1	Sequence 1, Appl
C 44	30.2	2.7	4692	2	US-08-972-631-1	Sequence 1, Appl
C 45	30.2	2.7	4692	2	US-08-972-629-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-920-812-17/c
; Sequence 17, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
US-08-920-812-17


```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921.177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
; US-08-921-177-17

```

```

Query Match      8.2%; Score 91; DB 1; Length 3615;
Best Local Similarity 53.7%; Pred. No. 6.4e-19;
Matches 240; Conservative 0; Mismatches 195; Indels 12; Gaps 2;

QY 610 atccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcagtcacatggtt 669
DB 1798 ATCTGTTACCGCATGACGGGAGAGATTGTCGCTCATTTCACGTCGCTGATACACGTTT 1739

QY 670 gccagcgaaagcttgctcggttgccatcgctcccaacaaagggcagttagttct 729
DB 1738 GCCAGTGGCGCTGTGGGTAAGGTAATTGCGATTTCGCCCTCGGTTGGTGAAGTGCCTTCT 1679

QY 730 ccggtgagtgaaagattgtggtggcattcccatctggccatgcttccagttccacc 789
DB 1678 CCGGTTCGGGTCGAATTCTTTCCGCTCAGCTCAACGTTGGTTCGCCACATTACA-----CGCCATTGGCATT 1625

QY 790 aaggctgagtggtttcccaatgtgatatcttgatgcacattggtttcgacacagtaaac 849
DB 1624 GAGTCAGATGATGG-----TGTGGAGATCCTTGATTCATGTCGATATCGACACCGTAAAA 1571

QY 850 ctcaacggcagcactttaaccgcgtgaagaagcagggcgatgaagtcacagggag 909
DB 1570 CTGGACGGCAAAATCTTTTCCGCTCAGCTCAACGTTGGTTCGCCACATTACA-----CGCCATTGGCATT 1511

QY 910 ctgctgtgtaattcgatattgatccattaaaggtgcaggttatgaggttaaccacgc 969
DB 1510 CGGCTGATTTCTTTTGATATCCCTGCTATTCGAGGCGCGGATTTGATCTGACGACGCG 1451

QY 970 attgtgtttcgaattacaagaaacccgacctgtaaacactacgtttgtggcgaaatt 1029
DB 1450 GTATTAAATCAGTAATAGCGATGATTTTACGGAGCGTATTATACCCACGCGCGGACGATA 1391

QY 1030 gaagcggagcacaacctgtcaactc 1056
DB 1390 AGCGCAGGTGAACCGCTGTATTCCATC 1364

RESULT 4
US-08-362-577C-17/c
; Sequence 17, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:

```

```

; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,577C
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
; US-08-362-577C-17

```

```

Query Match      8.2%; Score 91; DB 1; Length 3615;
Best Local Similarity 53.7%; Pred. No. 6.4e-19;
Matches 240; Conservative 0; Mismatches 195; Indels 12; Gaps 2;

QY 610 atccaggcacctttgaccggtgaagctattgcactgagcagcgtcagcagtcacatggtt 669
DB 1798 ATCTGTTACCGCATGACGGGAGAGATTGTCGCTCATTTCACGTCGCTGATACACGTTT 1739

QY 670 gccagcgaaagcttgctcggttgccatcgctcccaacaaagggcagttagttct 729
DB 1738 GCCAGTGGCGCTGTGGGTAAGGTAATTGCGATTTCGCCCTCGGTTGGTGAAGTGCCTTCT 1679

QY 730 ccggtgagtgaaagattgtggtggcattcccatctggccatgcttccagttccacc 789
DB 1678 CCGGTTCGGGTCGAATTCTTTCCGCTCAGCTCAACGTTGGTTCGCCACATTACA-----CGCCATTGGCATT 1625

QY 790 aaggctgagtggtttcccaatgtgatatcttgatgcacattggtttcgacacagtaaac 849
DB 1624 GAGTCAGATGATGG-----TGTGGAGATCCTTGATTCATGTCGATATCGACACCGTAAAA 1571

QY 850 ctcaacggcagcactttaaccgcgtgaagaagcagggcgatgaagtcacagggag 909
DB 1570 CTGGACGGCAAAATCTTTTCCGCTCAGCTCAACGTTGGTTCGCCACATTACA-----CGCCATTGGCATT 1511

QY 910 ctgctgtgtaattcgatattgatccattaaaggtgcaggttatgaggttaaccacgc 969
DB 1510 CGGCTGATTTCTTTTGATATCCCTGCTATTCGAGGCGCGGATTTGATCTGACGACGCG 1451

QY 970 attgtgtttcgaattacaagaaacccgacctgtaaacactacgtttgtggcgaaatt 1029

```


COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/117,121
FILING DATE: 20-NOV-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/CA97/00062
FILING DATE: 30-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 016252-001610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 292 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

Query Match 2.9%; Score 32.2; DB 1; Length 7218;

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: March 22, 2002, 08:53:35 ; Search time 157.11 Seconds
 (without alignments)
 6051.643 Million cell updates/sec

Title: US-09-604-231-3
 Perfect score: 1109
 Sequence: 1 tatgatttcggcggtccagt.....gttgaaacctgagttgctg 1109

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_1101.*
 1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
 2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
 3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
 4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
 5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
 6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
 7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
 8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
 9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
 10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
 11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
 12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
 13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
 14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
 15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
 16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
 17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
 18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
 19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
 20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
 21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
 22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1109	100.0	1109	22 AAF31529	C-glutamicum phosph
2	1107.4	99.9	1527	22 AAF31528	C-glutamicum phosph
3	1107.4	99.9	349980	22 AAH68533	C glutamicum codin
4	1084.4	97.8	1983	22 AAF67869	C glutamicum codin
5	1049.8	94.7	5969	22 AAF32543	Brevibacterium lac
6	128.6	11.6	5840	20 AAX12968	Enterococcus faeca
7	123.2	11.1	2913	22 AAH54100	S. epidermidis gen
8	118.8	10.7	3895	19 AAV52334	Streptococcus pneu
9	91	8.2	3615	15 AAQ55752	Escherichia coli g
10	90.2	8.1	9769	19 AAV52163	Streptococcus pneu
11	76.4	6.9	465	18 AAV75628	Staphylococcus aur

12	67	6.0	8494	19	AAV52296	Streptococcus pneu
13	66	6.0	1947	22	AAH81338	Escherichia coli p
c 14	65.2	5.9	9797	20	AAX13487	Enterococcus faeca
15	59.2	5.3	2049	22	AAH68426	C glutamicum codin
16	59.2	5.3	2172	22	AAF31542	C glutamicum codin
17	59.2	5.3	349980	22	AAH68528	C glutamicum codin
18	54.2	4.9	428	22	AAF31544	C glutamicum phosph
c 19	53.2	4.8	2996	22	AAH54445	S. epidermidis gen
c 20	53.2	4.8	3081	22	AAH54946	S. epidermidis gen
c 21	53.2	4.8	3932	22	AAH54056	S. epidermidis gen
c 22	51.8	4.7	7156	20	AAX12966	Enterococcus faeca
c 23	51.4	4.6	6285	20	AAX13352	Enterococcus faeca
c 24	51.4	4.6	2882	19	AAV52273	Streptococcus pneu
c 25	51	4.6	29555	18	AAV74517	Staphylococcus aur
c 26	49.2	4.4	796	18	AAV74736	Staphylococcus aur
c 27	46	4.1	1906	20	AAX13595	Enterococcus faeca
c 28	45	4.1	3604	20	AAX13417	Enterococcus faeca
c 29	44.6	4.0	7900	18	AAV74449	Staphylococcus aur
c 30	44.2	4.0	6092	20	AAX13162	Enterococcus faeca
c 31	43.2	3.9	760	19	AAZ96397	S. pneumoniae deri
c 32	43	3.9	29555	18	AAV74517	Staphylococcus aur
c 33	42.8	3.9	30246	18	AAV74367	Staphylococcus aur
c 34	42	3.8	400	18	AAV78285	Staphylococcus aur
c 35	41.6	3.8	341	18	AAV78546	Staphylococcus aur
c 36	40.6	3.7	567	21	AAA29550	HIV codon altered
c 37	40	3.6	249	20	AAX13801	Enterococcus faeca
c 38	39.8	3.6	465	18	AAV74502	Partial P. rubrum
c 39	39.4	3.6	3889	22	AAH54860	S. epidermidis gen
c 40	38.6	3.5	2365	18	AAV74406	Staphylococcus aur
c 41	37.6	3.4	10732	21	AAA10594	Gene encoding a su
c 42	37.4	3.4	5059	21	AAZ36227	DNA encoding a per
c 43	37.2	3.4	65632	21	AAH81502	N. meningitidis pa
c 44	37.2	3.4	349980	21	AAF21544	Neisseria meningit
45	37.2	3.4	1437668	21	AAH81490	N. meningitidis B

ALIGNMENTS

RESULT 1

AAF31529

ID AAF31529 standard; DNA; 1109 BP.

XX AAF31529;

XX 09-APR-2001 (first entry)

XX C.glutamicum phosphoenolpyruvate DNA #2.

XX Phosphoenolpyruvate; sugar phosphotransferase system; PTS; ds.

XX Corynebacterium glutamicum.

XX WO200102583-A2.

XX 11-JAN-2001.

XX 27-JUN-2000; 2000WO-IB00973.

XX 01-JUL-1999; 99US-0142691.

XX 23-AUG-1999; 99US-0150310.

XX 03-SEP-1999; 99DE-1042095.

XX 03-SEP-1999; 99DE-1042097.

XX (BADI) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-080989/09.

XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;

XX sugar phosphotransferase system proteins or their portions, useful for

XX typing or identifying C. glutamicum or related bacteria, and as markers


```
Query Match          99.9%; Score 1107.4; DB 22; Length 1527;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tatgatttcggtccagtcggtctgctctcggtctgctgtctgtctactaccacacgtcgc 60
Db 110909 TATGATTTCGGTGGTCCAGTCGGGCTGCTCTCGGCTGCTGTCTACTACCAATCGTC 110850

Qy 419 tatgatttcggtccagtcggtctgctctcggtctgctgtctgtctactaccacacgtcgc 478
Db 110909 TATGATTTCGGTGGTCCAGTCGGGCTGCTCTCGGCTGCTGTCTACTACCAATCGTC 110850

Qy 61 atcaatggtctgcacacagtcctccgcgaattgagctgagctgttttaaccaggtgga 120
Db 110909 ATCAATGGTCTGCACACAGTCCTCCGCGCAATTGAGCTGAGCTGTTTAAACAGGTGGA 110850

Qy 479 atcaatggtctgcacacagtcctccgcgaattgagctgagctgttttaaccaggtgga 538
Db 110909 ATCAATGGTCTGCACACAGTCCTCCGCGCAATTGAGCTGAGCTGTTTAAACAGGTGGA 110850

Qy 121 tcttcacatctgcgaacggcatctatgcttaataatcgcccaagggtgcggcatgtttggca 180
Db 110909 TCTTCACATCTGCGAACGGCATCTATGCTTAATAATCGCCCAAGGTGCGGCATGTTTGGCA 110850

Qy 539 tcttcacatctgcgaacggcatctatgcttaataatcgcccaagggtgcggcatgtttggca 598
Db 110909 TCTTCACATCTGCGAACGGCATCTATGCTTAATAATCGCCCAAGGTGCGGCATGTTTGGCA 110850

Qy 181 gtgtctctccgtgcgaagtgaaagctcaaggcccttgcaagtgcttcacaggtgtctcc 240
Db 110909 GTGTCTCTCCGTGCGAAGTGAAAGCTCAAGGCCCTTGCAAGTGCTTCACAGGTGTCTCC 110850

Qy 599 gtgtctctccgtgcgaagtgaaagctcaaggcccttgcaagtgcttcacaggtgtctcc 658
Db 110909 GTGTCTCTCCGTGCGAAGTGAAAGCTCAAGGCCCTTGCAAGTGCTTCACAGGTGTCTCC 110850

Qy 241 gctgttcttggtattacagagcgtcgatcttcgtgtgaaaccttcgcctgcctggcg 300
Db 110909 GCTGTCTCTGTTGATTACAGAGCGTCGATCTTCTGTTGAAACCTTCGCCTGCCTGGCG 110850

Qy 659 gctgttcttggtattacagagcgtcgatcttcgtgtgaaaccttcgcctgcctggcg 718
Db 110909 GCTGTCTCTGTTGATTACAGAGCGTCGATCTTCTGTTGAAACCTTCGCCTGCCTGGCG 110850

Qy 301 tcttcacatctgcgttacccgacagctatcggtgcgctttgattgcactctttaataatc 360
Db 110909 TCTTCACATCTGC GTTACCCGACAGCTATCGGTGCGCTTTGATTGCACCTCTTAAATAATC 110850

Qy 719 tcttcacatctgcgttacccgacagctatcggtgcgctttgattgcactctttaataatc 778
Db 110909 TCTTCACATCTGC GTTACCCGACAGCTATCGGTGCGCTTTGATTGCACCTCTTAAATAATC 110850

Qy 361 aaggcagttgctgtggcgctcaggttctcttggtgtgtgtttctattgatctccagat 420
Db 110909 AAGGCAGTTGCTGTGGCGCTCAGGTTCTCTTGGTGTGTGTCTTCTATTGATCTCCAGAT 110850

Qy 779 aaggcagttgctgtggcgctcaggttctcttggtgtgtgtttctattgatctccagat 838
Db 110909 AAGGCAGTTGCTGTGGCGCTCAGGTTCTCTTGGTGTGTGTCTTCTATTGATCTCCAGAT 110850

Qy 421 atggtcattgttctgtgtgagctgttacccttcttcatttcatttcatttcatttcatttc 480
Db 110909 ATGGTCATTGTCTGTGTGAGCTGTTACCCTTCTTCACTTCTTCACTTCTTCACTTCTTCACTTCT 110850

Qy 839 atggtcattgttctgtgtgagctgttacccttcttcatttcatttcatttcatttcatttc 898
Db 110909 ATGGTCATTGTCTGTGTGAGCTGTTACCCTTCTTCACTTCTTCACTTCTTCACTTCTTCACTTCT 110850

Qy 481 gcttatgctcttactgtgtgcgcgaacgcagcattgatccagatgcaaccgtgct 540
Db 110909 GCTTATGCTCTTACTGTGTGCGCGAACGCAGCATTGATCCAGATGCAACCGTGTCT 110850

Qy 899 gcttatgctcttactgtgtgcgcgaacgcagcattgatccagatgcaaccgtgct 958
Db 110909 GCTTATGCTCTTACTGTGTGCGCGAACGCAGCATTGATCCAGATGCAACCGTGTCT 110850

Qy 541 ccagtgctctgaggaacacccaagccgaagcagacccgcagaaatttcaaacgat 600
Db 110909 CCAGTGCTCTGAGGAACACCCAAGCCGAAGCAGACCCGCAGAAATTTCAAACGAT 110850

Qy 959 ccagtgctctgaggaacacccaagccgaagcagacccgcagaaatttcaaacgat 1018
Db 110909 CCAGTGCTCTGAGGAACACCCAAGCCGAAGCAGACCCGCAGAAATTTCAAACGAT 110850

Qy 601 tccacacatccaggaacaccccttgaccggtgaagctattgcaactgagcagcgtcagcat 660
Db 110909 TCCACACATCCAGGAACACCCCTTGACCGGTGAAGCTATTGCAACTGAGCAGCGTCAGCAT 110850

Qy 1019 tccacacatccaggaacaccccttgaccggtgaagctattgcaactgagcagcgtcagcat 1078
Db 110909 TCCACACATCCAGGAACACCCCTTGACCGGTGAAGCTATTGCAACTGAGCAGCGTCAGCAT 110850

Qy 661 gccatgttgcagcggaagcttgctgcgcggttgccatgctcccaaccaaggggcag 720
Db 110909 GCCATGTTGCAGCGGAAGCTTGCTGCGC GGTTCGCATGCTCCCAACCAAGGGGCAG 110850

Qy 1079 gccatgttgcagcggaagcttgctgcgcggttgccatgctcccaaccaaggggcag 1138
Db 110909 GCCATGTTGCAGCGGAAGCTTGCTGCGC GGTTCGCATGCTCCCAACCAAGGGGCAG 110850

Qy 721 ttagtcttcctggtgagtggaagattgtgtggtgcatctccatctggccatcttccga 780
Db 110909 TTAGTCTTCTCCTGGTGAGTGGAAGATTGTGTGGTG CATCTCCATCTGGCCATCTTCCGA 110850

Qy 1139 ttagtcttcctggtgagtggaagattgtgtggtgcatctccatctggccatcttccga 1198
Db 110909 TTAGTCTTCTCCTGGTGAGTGGAAGATTGTGTGGTG CATCTCCATCTGGCCATCTTCCGA 110850

Qy 781 gttgcacaaaggctgaggtggttccaatgtggatattttgatgcacattggtttcgac 840
Db 110909 GTTGCAAAAGGCTGAGGTGGTTCCAAATGTGGATATTTGATGCACATTGGTTTCGAC 110850

Qy 1199 gttgcacaaaggctgaggtggttccaatgtggatattttgatgcacattggtttcgac 1258
Db 110909 GTTGCAAAAGGCTGAGGTGGTTCCAAATGTGGATATTTGATGCACATTGGTTTCGAC 110850

Qy 841 acagttaacctcaacggcagcagcacttcaaccgctgaagaagcagggcgatgaagtcaaa 900
Db 110909 ACAGTTAACTCAACGGCAGCAGC ACTTCAACCGCTGAAGAAGCAGGGCGATGAAGTCAAA 110850

Qy 1259 acagttaacctcaacggcagcagcacttcaaccgctgaagaagcagggcgatgaagtcaaa 1318
Db 110909 ACAGTTAACTCAACGGCAGCAGC ACTTCAACCGCTGAAGAAGCAGGGCGATGAAGTCAAA 110850

Qy 901 gcaggggagctgctgtgtaattcgatattgatgccattaaaggctgcaggttataggga 960
Db 110909 GCAGGGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGGCTGCAGGTTATAGGGA 110850

Qy 1319 gcaggggagctgctgtgtaattcgatattgatgccattaaaggctgcaggttataggga 1378
Db 110909 GCAGGGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGGCTGCAGGTTATAGGGA 110850

Qy 961 accagccgattgtgttttcgaattacaagaaaacccgacctgtaaaaacacttacggttg 1020
Db 110909 ACCAGCCGATTGTGTTTTCGAATTACAAGAAAACCCGACCTGTAAAAACACTTACGGTTG 110850

Qy 1379 accagccgattgtgttttcgaattacaagaaaacccgacctgtaaaaacacttacggttg 1438
Db 110909 ACCAGCCGATTGTGTTTTCGAATTACAAGAAAACCCGACCTGTAAAAACACTTACGGTTG 110850
```

QY 61 atcaactggtgtgcaaccagctctcccccgaattgagctgagctgtttaaaccagggtgga 120
Db 110849 ATCACTGGTGTGACCAAGTCTCTCCGCCCAATTGAGCTGGAGCTGTTTAAACCAAGGTTGA 110790
QY 121 tcttcaatcttcgcaacgagcatctatggttaataatcagccagggtgcgcatgtttggca 180
Db 110789 TCCTTCACTCTCCGCAACGGCATCTATGGCTAAATATCCGCCAGGTCGGCATGTTTGGCA 110730
QY 181 gttgttctctcgcgcaagagtgaaagctcaaggcccttcaggtgcttcaggtgtctcc 240
Db 110729 GTGTTCTCTCGGCAAGAGTGAAGACTCAAGGGCCCTTCAGGTGCTTCAGGTGTCTCC 110670
QY 241 gctgttcttggtattacggagcctgcgattctcgggtgtaaccttcctcgcctgcctgcgcg 300
Db 110669 GCTGTTCTTGGTATTACGGAGCCTGCATCTCGGTGTGAACCTTCGCCCTGGCTGCGCG 110610
QY 301 tcttcaatcggtatcggtacgcgagcatatcggtggtgcgctttgattgacactctttaatc 360
Db 110609 TTCTTCACTCGGTATCGGTACCGAGCTATCGGTGGCGCTTTGATTGCACTCTTTAATATC 110550
QY 361 aaggcaattgctgtggcgctgcaggtttcttgggtgttttctattgattgctccagat 420
Db 110549 AAGCAGTTCGGTGTGGCGCTGCAGGTTCTTGGGTGTTGTTCTATTGATGCTCCAGAT 110490
QY 421 atggtcaatgttcttggtgtgcagttgttaccttcttcacgttcgcatcgcgcgcagcatt 480
Db 110489 ATGTCATGTTCTTGGTGTGTCAGTGTGTTACCTTCTTCATCGCATTCGGCGCAGCGATT 110430
QY 481 gcttatgaccttacttggtgtccgcaacgcgagcatatgacagatgcaacgcgtgct 540
Db 110429 GCTTATGGCCTTTACTTGGTTCGCCGACACGGCAGCATTTGATCCAGATGCAACCGTGTCT 110370
QY 541 ccagtcctcgcaggaacacaaagcgaagcagagcagcccgacagaaatttcaaacgat 600
Db 110369 CCAAGTGCCTGCAGACGACCAAGCGAGAGAGAGACACCCGACGAATTTTCAACGAT 110310
QY 601 tccaccatcatccagcgaacctttgaocgggtgaagctattgtcactgagcagcgtcagcgt 660
Db 110309 TCCACCATCATCCAGGCACCCCTTTGACCGGTGAAGCTATTGCACTGAGCAGCGTCAGCGAT 110250
QY 661 gccatgttcgacggaagagcttgctcggcggtgcacatgctcccaacaaaggggcag 720
Db 110249 GCCATGTTTCCCAAGCGAAGAGCTTGGCTCGGGCGTGGCCATGCTCCCAACCAAGGGCGAG 110190
QY 721 ttatgttctcgggtgagtggaagattgtgtggcattcccatctgccaatgctttcgca 780
Db 110189 TTAGTTTCTCCGCTGAGTGAAGAGATTGTTGGTGGCATTCCTCATCTGCCATGCTTTCGCA 110130
QY 781 gttcgcaccaagcgtgaggtggttcccaatgttgatatcttgatgcacattggtttcgac 840
Db 110129 GTTCGCACCAAGGCTGAGGATGTTCCAAATGTGGATATCTTGATGCACATTGTTTCGAC 110070
QY 841 acagtaaacctcaacgcagcgaactttaaccgcctgaagaagcagggcgatgaagcaaa 900
Db 110069 ACAGTAAACCTCAACGGCAGCGACTTTAAACCCGCTGAAGAAGCAGGGCGGATGAAGTCAA 110010
QY 901 gcaggggagctgctgtgtaattcgtattgctgcatcattgaagctgcaggttatgaggtta 960
Db 110009 GCAGGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAAGGCTGCAGGTTATGAGTTA 109950
QY 961 accacccgattgttcttcggaattacaagaaacacggacacctgtaaacacttacggtttg 1020
Db 109949 ACCACCCGATTGTTCTTTCGAAATTACAAGAAACCGGACCTGTAACACTTACGGTTTG 109890
QY 1021 ggcgaattgaacgagagaccactgctcaacgtgcgcaaaagaagcgggtgcagca 1080
Db 109889 GCGGAATTGAACGGGAGCGAACCTGCTCAACGTCGCAAAAGAAAGCGGTGCCAGCA 109830
QY 1081 acaccataagttaaacctttgagtgttcg 1109
Db 109829 ACACCATAGTTGAACACTTGAGTGTTCG 109801

RESULT 4
AAH67869
ID AAH67869 standard; DNA; 1983 BP.
XX
AC AAH67869;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2904.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG92650.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
PS Claim 8; SEQ ID NO: 2904; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1983 BP; 410 A; 508 C; 558 G; 507 T; 0 other;

Query Match 97.8%; Score 1084.4; DB 22; Length 1983;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1085; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tatgatttcgggtccagtcgggtctgctcttcggtctggttctactcaccaatcgtc 60
Db 898 tatgatttcgggtccagtcgggtctgctcttcggtctggttctactcaccaatcgtc 957
QY 61 atcaactggtgcacacagtcctcccgccaattgagctgagctgtttaaaccagggtgga 120
Db 958 atcaactggtgcacacagtcctcccgccaattgagctgagctgtttaaaccagggtgga 1017
QY 121 tcttcatcttcgcaacgcatctatgctaatatcgcagggtgcgcatgttttgca 180
Db 1018 tcttcatcttcgcaacgcatctatgctaatatcgcagggtgcgcatgttttgca 1077

```
QY 181 gtgttcttctggtcgaagagtgaaagctcaaggccttgacagtgcttcagatgtctcc 240
    |||
Db 1078 gtgttcttctggtcgaagagtgaaagctcaaggccttgacagtgcttcagatgtctcc 1137
QY 241 gctgttcttgggtattacgagcgtcgatcttcggtgtgaaccttcgcctgcgctggccg 300
    |||
Db 1138 gctgttcttgggtattacgagcgtcgatcttcggtgtgaaccttcgcctgcgctggccg 1197
QY 301 ttcttcaatcgggtatcggtaccgcagctatcggtggcgtttgattgcaactctttaatatc 360
    |||
Db 1198 ttcttcaatcgggtatcggtaccgcagctatcggtggcgtttgattgcaactctttaatatc 1257
QY 361 aaggcagttgcgttggcgctgcaggtttcttgggttcttgggttcttcttattgattcctcagat 420
    |||
Db 1258 aaggcagttgcgttggcgctgcaggtttcttgggttcttgggttcttcttattgattcctcagat 1317
QY 421 atggtcatgttcttgggtggtgcagttgttaccttcttcatcgtcattcggcgagcagatt 480
    |||
Db 1318 atggtcatgttcttgggtggtgcagttgttaccttcttcatcgtcattcggcgagcagatt 1377
QY 481 gttattggcctttacttgggttcgcccgaagcggcagcagattgatccagatgcaaccgtgtct 540
    |||
Db 1378 gttattggcctttacttgggttcgcccgaagcggcagcagattgatccagatgcaaccgtgtct 1437
QY 541 ccagtgcttcgaggaacgaccaaagccgaagcagagaagcaccgcagaaattttcaaacgat 600
    |||
Db 1438 ccagtgcttcgaggaacgaccaaagccgaagcagagaagcaccgcagaaattttcaaacgat 1497
QY 601 tcacatcatcatcagcgaacactttgaccggtgaagctattgcactgagcagcgtcacgcat 660
    |||
Db 1498 tcacatcatcatcagcgaacactttgaccggtgaagctattgcactgagcagcgtcacgcat 1557
QY 661 gcatgtttgcagcggaagcgttgcgtgcggcgttgcacatgtcccaaccgaagggcgag 720
    |||
Db 1558 gcatgtttgcagcggaagcgttgcgtgcggcgttgcacatgtcccaaccgaagggcgag 1617
QY 721 tiagtttctccggtgagtggaagattgtgtggtcattccattccatctgcccattgttcgca 780
    |||
Db 1618 tiagtttctccggtgagtggaagattgtgtggtcattccattccatctgcccattgttcgca 1677
QY 781 gttcgaccaaagctgagatggttccaatgttgatattcttgatgcacattgttttcgac 840
    |||
Db 1678 gttcgaccaaagctgagatggttccaatgttgatattcttgatgcacattgttttcgac 1737
QY 841 aagtaaaactcaacgcagcgaactttaaccgcgtgaagaagcagggcgatgaagtcaaa 900
    |||
Db 1738 acagtaaacctcaacgcagcgaactttaaccgcgtgaagaagcagggcgatgaagtcaaa 1797
QY 901 gcaggggagctgctgtgtaattcgtatattgattgccattaaagctgcaggttatgaggtta 960
    |||
Db 1798 gcaggggagctgctgtgtaattcgtatattgattgccattaaagctgcaggttatgaggtta 1857
QY 961 accacgccgattgtgttttcgaattacaagaacccggacctgtaaacacattacggtttg 1020
    |||
Db 1858 accacgccgattgtgttttcgaattacaagaacccggacctgtaaacacattacggtttg 1917
QY 1021 ggcgaataatgaagcgggagccaaacctgtctcaacgtctgcgaagaaagaagcgggtgccagca 1080
    |||
Db 1918 ggcgaataatgaagcgggagccaaacctgtctcaacgtctgcgaagaaagaagcgggtgccagca 1977
QY 1081 acacca 1086
Db 1978 acacca 1983
```

RESULT 5

AAF32543

ID AAF32543 standard; DNA; 5969 BP.

XX

AC AAF32543;

XX

DT 20-APR-2001 (first entry)

XX

```
DE Brevibacterium lactofermentum sucrose PTS enzyme II DNA SEQ ID NO:1.
XX
KW Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;
KW phosphoenolpyruvate:carbohydrate phosphotransferase system; glucose;
KW coryneform bacterium; phosphoenolpyruvate-sugar transport system; ds.
XX
OS Brevibacterium lactofermentum.
XX
FH Key Location/Qualifiers
FT CDS 3779..5764
FT /tag= a
FT /product= "sucrose PTS enzyme II"
XX
XX WO200102584-A1.
XX
PD 11-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-JP04348.
XX
XX 02-JUL-1999; 99JP-0189512.
PR (AJIN ) AJINOMOTO CO INC.
XX
XX PA
XX PI Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;
XX
XX WPI; 2001-138150/14.
DR P-PSDB; AAB69080.
XX
XX Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT II obtained by cassette ligation-mediated amplification of downstream
PT domain of coryneform bacterium sucrose gene, with sucrose-binding
PT activity .
XX
XX Claim 3; Page 22-29; 45pp; Japanese.
XX
XX The present sequence encodes the Brevibacterium lactofermentum sucrose
CC PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system or
CC phosphoenolpyruvate-sugar transport system) enzyme II, which has
CC sucrose-binding activity. A coryneform bacteria produced with the
CC sucrose PTS enzyme II gene can have more efficient sugar uptake, and
CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene
CC and it's disrupted gene, such as one without the sucrose PTS function,
CC can be used to produce new breeds of coryneform bacterial strains to
CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can
CC have improved amino-acid and nucleic acid productivity.
XX
XX Sequence 5969 BP; 1388 A; 1575 C; 1578 G; 1428 T; 0 other;
SQ
Query Match 94.7%; Score 1049.8; DB 22; Length 5969;
Best Local Similarity 96.7%; Pred. No. 8.7e-314;
Matches 1072; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 1 tatgatttcggcgtccagtcgcggtctgctctcttcggtctggttactcaccatcgctc 60
    |||
Db 4676 tatgatttcggcgtccagtcgcggtctgctctcttcggtctggttactcaccatcgctc 4735
QY 61 atcaactggctgcaccagctctcccgccaattgagctggagctgtttaaccagggtgga 120
    |||
Db 4736 atcaactggctgcaccagctctcccgccaattgagctggagctgtttaaccagggtgga 4795
QY 121 tccttcaatcttcgaacggcatctatggtcataatgccaccagggtgcggcatgttttgcca 180
    |||
Db 4796 tccttcaatcttcgaacggcatctatggtcataatgccaccagggtgcggcatgttttgcca 4855
QY 181 gtgttcttctggcgaagagtgaaagctcaaggccttcaggtgcttcaggtgtctcc 240
    |||
Db 4856 gtgttcttctggcgaagagtgaaagctcaaggccttcaggtgcttcaggtgtctcc 4915
QY 241 gctgttcttgggtattacgagcgtcgatcttcggtgtgaaccttcgcctgcgctggccg 300
    |||
Db 4916 gctgttcttgggtattacgagcgtcgatcttcggtgtgaaccttcgcctgcgctggccg 4975
```



```

Db 2379 tgattactggtgcaatggtgcccgaattgcttattatcatgctgttgccttaacgtttg 2438
Qy 500 ttgcccgaacgycagcaattgatccagatgcaacgcgtgctccagtgctgcaggaacga 559
Db 2439 tcttac-----gtttgaagatcaacctaatccagaaacagcaactgaaanaacag 2489
Qy 560 ccaagccgaagcagaagcaccgcagcaattttcaaacgattccacatcatccaggca- 618
Db 2490 aaactgataagatggtggccacctglaaaacgaatcaagaagacaaaatttttagcaa 2549
Qy 619 --cctttgaccggtgaagctattgcactgagcagcgtcagcgatgccaatggttgcacgcg 676
Db 2550 gtcaacttcaagtgaaattttaccgctagaaaagtaacaagaccctgttttgccttcag 2609
Qy 677 gaaagcttggtcgcggtggtgccatctgcccacaaagggcgagtagttcttcctcggtga 736
Db 2610 gtgctttaggaagaggtgttgcaattgagccgactgaagcgcaaacgtatgcaccgcag 2669
Qy 737 gtgaaagattggtggcattcccatctgcccattgcttccagtgcttgcagttcgaaccaagcgtg 796
Db 2670 atgtgaaatcaccacattattccgacaggaacatgctgttggcttgacgacaacagagg 2729
Qy 797 aggatggttccaatgtgatatcttgatgcacattggtttcgacacagtaaacctcaacg 856
Db 2730 gcgttgaa-----tattaatgcattatggcatggatgatacgggtcgaattagatg 2777
Qy 857 gcacgcactttaaccgcgtgaagacgagggcgatgaagtcacaaagcagggagctgctgt 916
Db 2778 gtaagagctttgaattatcagtgaaacaaagtgattctgttcaaaaagagagattgctag 2837
Qy 917 gtgaattcgattgatgcattaaagctgcaggttatgaggttaacacgcgcgattgtgt 976
Db 2838 ttacttttgatgctgctccattaaagaagctggttatccggtgattacacgcgattgtg 2897
Qy 977 ttctgaattacaa 989
Db 2898 taacgaatacga 2910

RESULT 7
AAH54100/c
ID AAH54100 standard; DNA; 2913 BP.
AC AAH54100;
XX
XX 03-SEP-2001 (first entry)
DE
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3464.
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
OS
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 1030-1031; 2188pp; English.
PS

```

```

XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 2913 BP; 1072 A; 480 C; 419 G; 942 T; 0 other;

```

```

Query Match 11.1%; Score 123.2; DB 22; Length 2913;
Best Local Similarity 59.0%; Pred. No. 1.7e-27;
Matches 256; Conservative 0; Mismatches 163; Indels 15; Gaps 2;
Qy 1 tatgattcgcgccagtcggcggtgctcttcggtctgctggtctactcaccacgcgc 60
Db 2274 TATGAATTGGTGGAGCTATTGGTGGACTTATATTTGGTTTATATATGCGCCATCGTC 2215
Qy 61 atcactgctgcacagtccttccgcgaattgagctg-----gagctgctt 108
Db 2214 ATTACAGGAATGCACCATAGCTTTATTGAGTTGAAACGACATTAATTGCTGATCGGACT 2155
Qy 109 aaccaggttgatccttcatttcgcaacgcatctatgctaatatcgccaggtgcg 168
Db 2154 AAACAGGTGGTTCATTATCTCCCAATCGCAACGATGTCAATATTGCACAAGGTGT 2095
Qy 169 gcatgtttggcagtgcttct--tctggcgagagtgaaagctcaaggccttcgaggt 225
Db 2094 GCAGCTTTAGCTGCATCTTTATCATTRAGCAATAATAAAATTAAGAGTGTGTCTCC 2035
Qy 226 gcttcaggtgctccgcgtgtcttcttggtattacggagcctgcgactctcggtggaacctt 285
Db 2034 GCGGCGGGTATTTTCAGCTTTACTAGGAATTACAGAACCAACCAATGTTGGTGTCAATCTT 1975
Qy 286 ggcctgcgtggcgttcttcctcgtatcggtaccgagctatcggtgcgcttgatt 345
Db 1974 AAATTGAGATATCCATTATAGTGCTGTGTCAGGATCAGGTATAGTGCAGCTTATAT 1915
Qy 346 gcactctttaatacgaagcaggtgctgtggcgctgcaggttcttctgggtgtgttct 405
Db 1914 TCATTCTTCAAAGTAAAGCGATAGCGCTTGTCACAGCTGGATTACCTGGATTATATCT 1855
Qy 406 attgatgtccaga 419
Db 1854 ATAAATCTACACA 1841

```

```

RESULT 8
AAV52334/c
ID AAV52334 standard; DNA; 3895 BP.
XX
XX AC AAV52334;
XX
XX 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:201.
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

```


[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 22, 2002, 08:49:30 ; Search time 2009.21 Seconds
(without alignments)
9105.754 Million cell updates/sec

Title: US-09-604-231-3
Perfect score: 1109
Sequence: 1 tatgatttcgctccagt.....gttgaacctgagtgttcg 1109

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vl.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.om.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.sy.*
28: em.un.*
29: em.vl.*
30: em.htgo_hum.*
31: em.htgo_inv.*
32: em.htgo_rod.*
33: em.htg_hum.*
34: em.htg_inv.*
35: em.htg_rod.*
36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1109	100.0	1109	6	AX069136	AX069136 Sequence
2	1107.4	99.9	1527	6	AX069134	AX069134 Sequence
c 3	1107.4	99.9	349980	6	AX127152	AX127152 Sequence
4	1084.4	97.8	1983	6	AX122988	AX122988 Sequence
5	165.6	14.9	2508	1	STRSCRA	M2711 Streptococ
c 6	138.4	12.5	21838	1	PDCRAFOPER	L32093 Pedococcus
c 7	138.4	12.5	21838	1	PPSURFOP	Z32771 P.pentosa
8	135.2	12.5	296950	1	AP001508	AP001508 Bacillus
9	135.2	12.2	2655	1	SXSCRA	X69800 S.xylosus s
c 10	133	12.0	5800	1	LLZ97015	Z97015 Lactococcus
c 11	127	11.5	10264	1	AE004395	AE004395 Vibrio ch
c 12	126.6	11.4	10085	1	AE006222	AE006222 Pasteurel
c 13	123.2	11.1	2913	1	AF269422	AF269422 Staphyloc
c 14	123.2	11.1	2913	6	AX144742	AX144742 Sequence
c 15	121	10.9	341350	1	AP003365	AP003365 Staphyloc
c 16	120.4	10.9	151947	2	SPNEU1902	AL449924 Streptoco
17	120	10.8	2955	1	VBSCRAK	M76768 Vibrio algi
18	119.4	10.8	4158	1	BSTREAPR	Z54245 B.subtilis
c 19	118.8	10.7	14515	1	AE007479	AE007479 Streptoco
c 20	115.8	10.4	22197	1	D83967	D83967 Bacillus su
c 21	115.8	10.4	37900	1	D86417	D86417 Bacillus su
c 22	115.8	10.4	208430	1	BSUB0005	Z99108 Bacillus su
c 23	110.4	10.0	299850	1	AP001514	AP001514 Bacillus
24	104.8	9.4	3518	1	BACSACP	J03006 B.subtilis
25	104.8	9.4	97015	1	BSGENR	X73124 B.subtilis
c 26	104.8	9.4	212150	1	BSUB0020	Z99123 Bacillus su
27	104.4	9.4	4487	1	AF229829	AF229829 Pseudomon
28	101.4	9.1	5065	1	ERWBGPA	M1772 Erwinia chr
c 29	100.6	9.1	2819	1	BACISPO	D37921 Alkalophilol
30	98.4	8.9	8006	1	AF206272	AF206272 Streptoco
31	97.8	8.8	2543	1	BSTREPA	X80203 B.subtilis
c 32	97.8	8.8	292550	1	AP001513	AP001513 Bacillus
c 33	92.8	8.4	12086	1	AE006630	AE006630 Streptoco
c 34	92.2	8.3	136254	1	ECOW82	L10328 E. coli; th
35	91	8.2	2134	1	ECOBGLC	M15746 E.coli bgIC
c 36	91	8.2	3615	6	AR012068	AR012068 Sequence
c 37	91	8.2	3615	6	AR025193	AR025193 Sequence
c 38	91	8.2	3615	6	AR038650	AR038650 Sequence
c 39	91	8.2	3615	6	AR068497	AR068497 Sequence
40	91	8.2	5270	1	ECOBGLQ	M16487 E.coli bgl
c 41	91	8.2	12989	1	AE000449	AE000449 Escherich
c 42	90.2	8.1	10371	1	AE007465	AE007465 Streptoco
43	88.6	8.0	2343	3	AF309637	AF309637 Leishmani
44	86.6	7.8	300750	1	AP001509	AP001509 Bacillus
45	85.8	7.7	4621	1	KPSCRYAB	X57401 K.pneumonia

ALIGNMENTS

RESULT 1

AX069136
LOCUS AX069136 1109 bp DNA
DEFINITION Sequence 3 from Patent WO0102583.
ACCESSION AX069136
VERSION AX069136.1 GI:12579018
KEYWORDS
SOURCE
ORGANISM

Corynebacterium glutamicum.
Corynebacterium glutamicum.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.

REFERENCE 1 (bases 1 to 1109)

AUTHORS Pompeius M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.

TITLE Corynebacterium glutamicum genes encoding phosphoenolpyruvat e;

JOURNAL sugar phosphotransferase system proteins

Patent: WO 0102583-A 3 11-JAN-2001;

BASES ARTIENGESCHAF (DE)

FEATURES Location/Qualifiers

1..1109

source

/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"

CDS

<1..1089
/note="FRXA00315"
/codon_start=1
/transl_table=11
/protein_id="CAC27277.1"
/db_xref="GI:12579019"

/translation="YDFGPPVGGLLFLVSPVITGLHOSPPPIELFLNOGGSFIF
ATASMANIAGGACLAFLAKSEKLGLAGASVSAYVIGITEPAIFGVNLRWRWFF
IGTGAIGALIALNRIKAVLGAAGFLGVSIDAPDMVMFLVCAVTTFFIAFGAAI
AYGLIVRRNSIDPATAAPVPAGTAKAEAPAEFSNDSTIIQAPLTGEAIALSSV
SDAMFASGLSGVAIVPTKQLVSPVSKIVVAFPSHAFVRTKAEDGSNDVILMH
IGFDVNLNTHFNPULKOGDEVKAGELLCEFDIDAIIKAAGYEVTPPIVVSNNKKTGP
VNTYIGLGETEAGANLLNVAKKEAVPATP"

BASE COUNT 236 a 278 c 303 g 292 t
ORIGIN

Query Match 100.0%; Score 1109; DB 6; Length 1109;
Best Local Similarity 100.0%; Pred. No. 1.4e-297;
Matches 1109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tatgatttcggcggtccagtcggtctctctctctctctctctctctctctctcaccacatcgc 60
Db 1 TATGATTTCGGCGGTCCAGTCGGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 60
QY 61 atcaactggtctgcaacagctctcccgccaattgagctgagctggttttaaccaggtgga 120
Db 61 ATCACTGGTCTGCACCAAGTCTCTCCCGCCAATTGAGCTGGAGCTGTTAACCAAGGTGGA 120
QY 121 tcttcaatcttcgcaacggtcatctatgcttaatatcgccagggtggcgcatgtttggca 180
Db 121 TCCTTCATCTTCGCAACGGCATCTATGCTTAATATCCCGCAGGTGGCGCATGTTTGCCA 180
QY 181 gtgttcttcctgcaagagtgaaagctcaaggcccttcagggtgttcagggtgtctcc 240
Db 181 GTGTCTTCCTGGCGAAGAGTGAAAGCTCAAGGGCCCTTCAGGTGCTTCAGGTGTCTCC 240
QY 241 qctgtcttgattacgagcctcgatcttcggtgagctgagctggaaccttcgctgcgtgccc 300
Db 241 GCGTGCTTGGTATACGGAGCTCGCATCTTCGGTGTGAACCTTCGCCGTGCGGTGCCG 300
QY 301 tcttcaatcggtatcgtaacgagctatcggtggtggtggtggtggtggtggtggtggtg 360
Db 301 TTCTTCATCGTATCGTACCGCAGCTATCGGTGGCGCTTTGATTGACCTTTTAATATC 360
QY 361 aaggcagttgcttggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 420
Db 361 AAGGCAGTTGCGTGGCGCTGCGAGTTTCTTGGGTGTTCTTGGTGTGTTCTATTGATGCTCCAGAT 420
QY 421 atggtcatgttcttggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 480
Db 421 ATGTCATGTTCTGGTGTGCGAGTTGTTTACCTTCTTCATCGATTCGCGCGAGCGGATT 480
QY 481 gcttatggccttacttggttcgcgcgaacgagcagcagcagcagcagcagcagcagcagc 540
Db 481 GCTTATGGCTTTACTTGGTTCGCGCAGCAGCGCAGCATTTGATCCAGATGCAACCGTGCT 540
QY 541 ccagtgcttcaggaagcagcaagcagcaagcagcaagcagcaagcagcaagcagcaagcag 600
Db 541 CCAAGTGCCTGCAGGAACGACCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 tccaccatcatcaggacaccttgacggtggaagctatgctgacgtgacgtgacgtgacgtg 660
Db 601 TCCACCATCATCCAGGACCTTTGACCGGTGAAGCTATTGCACTGACGACGCTCAGCGAT 660
QY 661 gccatgttgccagcggaagcttggtcgtggcggttggtcgtcgtcgtcgtcgtcgtcgtcgt 720
Db 661 GCATGTTTGCCAGCGGAAGCTTGCTCGGGGCTTGCCATCTGCCACCAAGGGCGAG 720
QY 721 ttatgttctccggtgagtggaagatgtgtgtggcatttccattccattcgtggtcgtcgtcgt 780
Db 721 TTAGTGTCTCCGCTGAGTGGAAAGATGTGTGTGGCATTCCCATCTGCTCCTCCTCCTCCTCCT 780

Db 721 TTAGTGTCTCCGCTGAGTGGAAAGATGTGTGTGCTCATCCATCTGCCCATGCTTTGCGCA 780
QY 781 qticgcaccaagctgaggtggttccaatgtgatatcttgatgcacattggtttcgac 840
Db 781 GTTCGCACCAAGGCTGAGGATGGTTCATATGTGGATATCTTGTATGTCACATGTTGTTCCGAC 840
QY 841 acagtaaacctcaacggcgcacgcactttaaccgcctgaagaagcaggcgatgaagacaaa 900
Db 841 ACAGTAACCTACACGCGCAGCACTTTAACCCGCTGAAGAAGCAGGCGGATGAAGTCAA 900
QY 901 gcaggggagctgctgtgtaattcgatctatgctccattagtcaggtcaggttatgaggtg 960
Db 901 GCAGGGAGCTGCTGTGTAATTCATATTCATGCCATTAAAGGCTGAGGTTATGAGTA 960
QY 961 accacgctgattgtttcgaattacaagaacacccgacctgtaaacacttaccggtttg 1020
Db 961 ACCACGCGGATTGTTTTCGAATTACAAGAAACCGGACCTGTAAACACTTACGGTTTG 1020
QY 1021 ggcgaattgaagcgggagcgaacctgctcaacgtctcaacgtcgaagaagcagcggtgccagca 1080
Db 1021 GCGGAATTGAAGCGGCGGAGCAACCTGCTCAACGCTGCAAGAAGAAAGCGGTGCCAGCA 1080
QY 1081 acaccataagttgaaccttgagtggttcg 1109
Db 1081 ACACCATAAAGTTGAACCTTTGAGTGTTCG 1109

RESULT 2

AX069134 1527 bp DNA PAT 25-JAN-2001
LOCUS Sequence 1 from Patent WO0102583.
DEFINITION AX069134
ACCESSION AX069134
VERSION AX069134.1 GI:12579016
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
Corynebacterium.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..1527

/organism="Corynebacterium glutamicum"

/db_xref="taxon:1718"

101..1507

/note="RXS00315"

/codon_start=1

/transl_table=11

/protein_id="CAC27276.1"

/db_xref="GI:12579017"

/translation="NAMVPSLVNGYDVAATVAAGBPMWSLFLDVAOAGYOGTVLP

VLVVSWILATIEKFLHKLKGTADFLITVLLTLLTGLTFTAGIPAMRWGVDVLAHG

LOGLDFFGPGVGLFLVSPVITGLHOSPPPIELFLNOGGSFIFATASMANIAQG

GAACLAFFFLAKSEKLGLAGASVSAYVIGITEPAIFGVNLRWRWFFIGTAAIAGG

ALIALFNRIKAVLGAAGFLGVSIDAPDMVMFLVCAVTTFFIAFGAAIAYGLIVRRN

GSIDPDATAAPVPAGTAKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKL

GGSAIVPTKQLVSPVSKIVVAFPSHAFVRTKAEDGSNDVILMHIGFDVNLNG

THENPLKOGDEVKAGELLCEFDIDAIIKAAGYEVTPPIVVSNNKKTGPVNTYGLGIE

ACANLLNVAKKEAVPATP"

BASE COUNT 304 a 392 c 430 g 401 t

ORIGIN

Query Match 99.9%; Score 1107.4; DB 6; Length 1527;

Best Local Similarity 99.9%; Pred. No. 4.1e-297;

Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tatgatttcggcggtccagtcggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 60

```
|||||
Db 419 TATGATTTCCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTCTGCTACTACCAATCGTC 478
QY 61 atcaactgtctgcacccagtccttccgcgaattgagctggagctgttttaaccagggtgga 120
|||||
Db 479 ATCACTGGTCTGCACCAAGTCTTCGCCCAATAGAGCTGAGCTGTTAAACAGGGTGGA 538
QY 121 tcttcacatctgcgaacgcatctatggttaataatgcgccagggtgcgcgatgtttggca 180
|||||
Db 539 TCTCTCATCTTCGCAACGGCATCTATFGGCTAATATGCCCCAGGGTGCGCATGTTTGGCA 598
QY 181 gtgttcttctgcgaagagtgtaaaagctcaagggccttgcaaggcttgcaggtgtctcc 240
|||||
Db 599 GTGTTCTTCTGCGAAGAGTGAAGAGCTCAAGGGCTTCAGGGTCTTCAGGTGTCTCC 658
QY 241 gctgttcttggattacagagcctgcgatctcttgggtgtgtaaaccttcgcctgcgtggccg 300
|||||
Db 659 GCTGTGTTCTTGGTATTACGGAGCGCTCGGATCTTCGGTGTGAACCTTCGCCCTGGCGG 718
QY 301 tcttcacatcggtatcggtaccgcagctatcggtgcggtttgattgacatctttaatac 360
|||||
Db 719 TTTCTTCATCGGTATCGGTACCGAGCTATCGGGTGGCGCTTTGATTGCATCTCTTAATATC 778
QY 361 aaggcagttgcgttgggcgctgcaggtttcttgggtgtgttcttctattgatgctccagat 420
|||||
Db 779 AAGCGAGTTGCGTTGGCGCGCTGCAGGTTCTTGGGTGTGTTTCTATTATGCTCCAGAT 838
QY 421 atggtcatgttcttgggtgtgtgcagttgttaaccttcttcacatcgcatcggcgagcatt 480
|||||
Db 839 ATGGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATT 898
QY 481 gcttatggccttactgtttgcgcgaagcgcagcattgatccagatgaacccgtgct 540
|||||
Db 899 GCTTATGGCGCTTACTTGTGTTGCCCAACGGCAGCATTCGATTCAGATGCACCGCTGTG 958
QY 541 ccagtgctctgcaggaaacgcaaacgcaagcagaaagccgcagaaatcttcaaacgat 600
|||||
Db 959 CCAGTGCCTGCAGGAACGACCAAGCCGAAGCAGACAGCACCCGCAGAAATTCAAACGAT 1018
QY 601 tcacacatcatccaggcacctttgaacggtgaagctattgcaactgagcagcgtcacgcat 660
|||||
Db 1019 TCACACATCATCCAGCACCTTTGACCGGTGAAGCTATTGTCATGACGACGCTCAGCGAT 1078
QY 661 gcatgtttgcagcggaagcttgctgcgcgttgccatgctcccaaccgaagggcag 720
|||||
Db 1079 GCCATGTTTGCACGCGGAAGCTTGCTCGGGGTTGGCCATCTGCCAACCAAGGGGCGAG 1138
QY 721 ttagtcttccggtgagtggaagattgtgttggtgcatctcccatctggccatgctttcgca 780
|||||
Db 1139 TTAGTTTCTCCGTGAGTGGAAAGATTGTGGTGGCATTCCTCATCTGGCCATGCTTTCCGA 1198
QY 781 gttcgcaacaaaggctgaggatggttccaatgtggatatcttattgatgcacattggtttcgac 840
|||||
Db 1199 GTTCGCACCAAGGCTGAGGATGTTTCCAATGTGGATATCTTGATGCACATTTGGTTTCGAC 1258
QY 841 acagtaaacctcaacggcgcacactttaaccgctgaagaagcaggcgcatgaagtcaca 900
|||||
Db 1259 ACAGTAACCTCAACGGCAGCACCTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAA 1318
QY 901 gcaggggagctgctgtgtaattcgatatgtatgacctaaagcgtgaagcaggttatgaggtta 960
|||||
Db 1319 GCAGGGGAGCTGCTGTGTAATTGCAATATTCGATATTCGATGCAATTAAGGCTGCAGGTTATGAGGTA 1378
QY 961 accagcgcatgtgtgttctcgaattacaagaaaacccggacctgtaaacactttacggtttg 1020
|||||
Db 1379 ACCACGCCGATTTGTTTTCGAATTAACAAGAAAACCGGACCTGTAAACACTTTACGGTTTG 1438
QY 1021 ggcgaattgaagcgggaacacactgtccaacgtcgcaaaagaaagcagcgtgccaagca 1080
|||||
Db 1439 GCGCAAAATGAAGCGGGAGCCCAACCTGCTCAACGTGCGCAAAAGAAAGACGGGTGCCAGCA 1498
QY 1081 acaccataagttgaacaccttgaggttgcg 1109
|||||
```

```
Db 1499 ACACCATAAGTTGAAACCTTGAGTGTTCG 1527

RESULT 3
AX127152/c
LOCUS AX127152 349980 bp DNA PAT 11-MAY-2001
DEFINITION Sequence 7068 from Patent EP1108790.
ACCESSION AX127152 AX114121
VERSION AX127152.1 GI:14041140
KEYWORDS
SOURCE
  Corynebacterium glutamicum.
  Corynebacterium glutamicum
  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Corynebacterineae; Corynebacteriaceae;
  Corynebacterium.
REFERENCE 1 (bases 1 to 349980)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 7068 20-JUN-2001;
KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
  Location/Qualifiers
    1..349980
     /organism="Corynebacterium glutamicum"
     /db_xref="taxon:1718"
     /note="Seq 1 to long (3.309.400) split in 11, seq 7068
       2.700.001 3.049.980"
BASE COUNT 81250 a 97718 c 90621 g 80391 t
ORIGIN

Query Match 99.9%; Score 1107.4; DB 6; Length 349980;
Best Local Similarity 99.9%; Pred. No. 6e-297;
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tatgatttcgggcggctcagtcggcggtctgctcttcgggtgtgtctactacccaatcgctc 60
|||||
Db 110909 TATGATTTCCGGTGGTCCAGTCGGCGGTCTGCTCTTCGGTGTGCTACTACCAATCGTC 110850
QY 61 atcaactgtctgcacccagtccttccgcgaattgagctggagctgttttaaccagggtgga 120
|||||
Db 110849 ATCACTGGTCTGCACCAAGCTCTTCGCCCAATTAAGCTGAGCTGTTAAACAGGGTGGA 110790
QY 121 tcttcacatctgcgaacgcatctatggttaataatgcgccagggtgcgcgatgtttggca 180
|||||
Db 110789 TCTTCATCTTCGCAACGGCATCTATFGGCTAATATGCCCCAGGGTGCGCATGTTTGGCA 110730
QY 181 gtgttcttctgcgaagagtgtaaaagctcaagggccttgcaaggcttgcaggtgtctcc 240
|||||
Db 110729 GTGTTCTTCTCCGGAAGAGTGAAGAGCTCAAGGGCTTCAGAGTGTCTCAGGTGTCTCC 110670
QY 241 gctgttcttggattacagagcctgcgatcttctgggtgtgtaaaccttcgcctgcgtggccg 300
|||||
Db 110669 GCTGTGTTTGGTATTACGGAGCGCTGCGATCTTCGGTGTGAACCTTCGCCCTGGCGGCG 110610
QY 301 tcttcacatcggtatcggtaccgcagctatcggtgcgcgtttgattgacatctttaatac 360
|||||
Db 110609 TTTCTTCATCGGTATCGGTACCGCAGCTATCGGGTGGCGCTTTGATTGCATCTTTTAATATC 110550
QY 361 aaggcagttgcgttgggcgctgcaggtttcttgggtgtgttcttctattgatgctccagat 420
|||||
Db 110549 AAGCGAGTTGCGTTGGCGCGCTGCAGGTTCTTGGGTGTGTTTCTATTATGATCTCCAGAT 110490
QY 421 atggtcatgttcttgggtgtgtgcagttgttaaccttcttcacatcgcatcggcgagcatt 480
|||||
Db 110489 ATGGTCATGTTCTTGGTGTGTGCAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATT 110430
QY 481 gcttatggccttactgtttgcgcgaagcgcagcattgatccagatgaacccgtgct 540
|||||
Db 110429 GCTTATGGCGCTTACTTGTGTTGCCCAACGGCAGCATTCGATTCAGATGCACCGCTGTG 110370
QY 541 ccagtgctctgcaggaaacgcaaacgcaagcagaaagcagcagcagcagattttcaaacgat 600
|||||
```


Db 110369 CCACTGCTGCAGGAACAGCAAAAGCCGAAGACAGAACACCCGAGAAATTTTCAAAACGAT 110310
QY
601 tccaccatcatccaggccacctttgaacggtgaagctattgactgagcagctcagcgat 660
Db 110309 TCCACCATCATCCAGGACACCTTTGACCGGTGAAGCTATTGCACTGACAGCGTCAGCGAT 110250
QY
661 gcaatgtttccagcggaagagcttggctcggggttgcctcgtcccaaaccaaggggcag 720
Db 110249 GCCATGTTTGCACGCGAAGAGCTGGCTCGGCGCTTGCCATCGTCCCAACCAAGGGCGAG 110190
QY
721 ttaattctccggtgagtggaagattggtggcattcccatgtgcccattgctttcgca 780
Db 110189 TTAGTTTTCTCCGGTGAAGTGAAGAAATGTTGGTGGCAATGCCATGTGCCATGCTTTGCGCA 110130
QY
781 gttccaccaggtgagtggttccaatgtgatatcttgatgacattggtttcgac 840
Db 110129 GTTCGCACCAAGGCTGAGGATGGTTCCAAATGTGGATATCTTGATGCAATGGTTTTCGAC 110070
QY
841 acagtaaacctcaacggcagcgactttaaccctgtaagcagcagggcgatgaagtcaaa 900
Db 110069 ACAGTAAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGCGCATGAAGTCAAA 110010
QY
901 gcaggagctgctgtgtaattgatgattgatgacattgaagctgaggttatgaggtta 960
Db 110009 GCAGGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAGGCTGCAGGTTATGAGGTA 109950
QY
961 accagccgattgttttcgaattcaacgaacacggacgttaaacacttacggtttg 1020
Db 109949 ACCACGGCGATGTGTTTGAATTAACGAATAACCGACCTGTAAACACTTACGGGTTG 109890
QY
1021 ggcgaattgaagcggagcgaacctgtcaacgtcgcaagaaagcgggtgccagca 1080
Db 109889 GCGCAATTGAGGGGAGCCACCTGCTCAACGTCCCAAGAAAGCGGTGCCAGCA 109830
QY
1081 acaccataagtgaaacctgagtggttcg 1109
Db 109829 ACACCATAAGTTGAACCTTGAGTGTTCG 109801

RESULT 4

AX122988 1983 bp DNA PAT 11-MAY-2001
LOCUS
DEFINITION Sequence 2904 from Patent EP1108790.
ACCESSION AX122988
VERSION AX122988.1 GI:14040476
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 1983)
Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
Novel polynucleotides
Patent: EP 1108790-A 2904 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source
1..1983
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 410 a 508 c 558 g 507 t
ORIGIN

Query Match 97.8%; Score 1084.4; DB 6; Length 1983;
Best Local Similarity 99.9%; Pred. No. 1e-290;
Matches 1085; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 tatgatttcggcgctcagtcggcggtctgctctcttcggtctgcttactcaacatgctc 60
Db 898 TATGATTTCCGGTGTCCAGTCGGCGGCTGCTCTTCCGGTGTGCTACTCACCAATGCTC 957

QY 61 atcactgggtctgcaccagtccttcccaccattgagctgagctgtttaaccagggtgga 120
Db 958 ATCACTGGTCTGCACCAAGTCTCTTCCGCCCAATTGAGCTGGAGCTGTTTAACCAAGSGTGA 1017
QY 121 tcttcaatcttcgcaacgagcatctatgcttaataatcgcacagggtgaggatgtttggca 180
Db 1018 TCTTTCATCTTCCGACACGGCATCTATGGCTAATATCCGCCAGGGTGGCATGTTTGGCA 1077
QY 181 gttcttctctcgcgaagagtgaaagctcaaggcccttcaggtgcttcaggtgtctcc 240
Db 1078 GTGTTCTTCCGCGGAAGAGCTGAAGAGCTCAAGGCCCTTCAGAGGTGTTTCAGGTGTCTCC 1137
QY 241 gctgttcttggtattaccggagcctgcgactctcggtgtgaaccttcgctcgcgtgcgcg 300
Db 1138 GCTGTTCTTGGTATTACGGAGCCTGCGATCTTCGGTGTGAACCTTTCGCTGCGCTGCGCG 1197
QY 301 tcttcaatcggatcgggtacgcagctatcgggtggcgcttgcacactctttaataatc 360
Db 1198 TTTCTTCATCGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGACACTCTTTAAATATC 1257
QY 361 aaggcagttcgggtggcgctgcaggttcttggtgtattttctattgattgctccagat 420
Db 1258 AAGCAGTTCGGTGGCGCTGCAGGTTCTTGGGTGTTGTTCTATTGATGCTCCAGAT 1317
QY 421 atggtcatgttcttggtgtgcagttgttaccttcttcattcgcattcgcgcgcgcgatt 480
Db 1318 ATGGTCATGTTCTGGTGTGTCAGTGTGTTACCTTCTTCATCGCATTCGCGCAGCGCAT 1377
QY 481 gcttatggccttacttggttcggcgcaacggcagcattgatccagatgcacacgcgtgct 540
Db 1378 GCTTATGGCTTTACTTGGTTCGCCGCAACGGCAGCATTTGATCCAGATGCAACCGCTGCT 1437
QY 541 ccagtgcttcaaggaacacaaaagccgaagcagaagacccgcagaaattttcaaacgat 600
Db 1438 CCAAGTGCCTGCAGGACAGCAAAAGCCGAAGACAGACACCCGACAGAAATTTTCAAACGAT 1497
QY 601 tcaccateatccaggcacctttgaccggtgaagctattgcactgagcgcgtcagcgat 660
Db 1498 TCCACCATCATCCAGCACCCTTTGACCGGTGAAGCTATTGCATGACGACGGTCAGCGAT 1557
QY 661 gccatgtttccagcggaagccttgctcggcggttgccatcttgcacacattggtttcgac 720
Db 1558 GCCATGTTTCCAGCGGAAAAGCTTGGCTCGGCGCTTGCCATCGTCCCAACCAAGGGCGAG 1617
QY 721 ttaatttccggtgagtgaaagattggtggcattcccatctggcagatgctttcgca 780
Db 1618 TTAGTTTCTCCGGTGAAGTGAAGATTGTTGGTGGCATTTCCCATCTGCGCCATGCTTTGCGCA 1677
QY 781 gttccaccaagctgaggtggttccaatgtggatatcttgatgcacattggtttcgac 840
Db 1678 GTTCGCACCAAGGCTGAGGATGGTTCCAATGTGGATATCTTGATGCACATTGGTTTCGAC 1737
QY 841 acagtaaacctcaacggcagcactttaaccctgaagcagcaggcgatgaagtcaaa 900
Db 1738 ACAGTAAACCTCAACGGCAGCGACCTTTAAACCGCTGAAGAAGCAGGCGCATGAAGTCAAA 1797
QY 901 qcaggagctgctgtgtgaattcgattgatgacattgaagctgaagctgaggttatgagta 960
Db 1798 GCAGGGAGCTGCTGTGTGAATTCGATATTGATGCCAATTAAAGCTGCAGGTTATGAGGTA 1857
QY 961 accagccgattgttttcgaattcaagaacacggacgtgtaaacacttacggtttg 1020
Db 1858 ACCAGCCGATGTTGTTTCGAATTACAGAAACCGGACCTGTAAACACTTACGGGTTG 1917
QY 1021 ggcgaattgaagcgggagccaaacctgctcaacgctgcgaagcgaagcgaagcgggtgccagca 1080
Db 1918 GCGCAAAATTAAGCGGAGCGCAACCTGCTCAACGTCGCAAAAGAAAGCGGTGCCAGCA 1977
QY 1081 acacca 1086
Db 1978 ACACCA 1983

DNA; Insertion sequence IS3 homolog (transposable element Insertion sequence IS3 homolog, kingdom Prokaryotae) DNA; and *Pediococcus pentosaceus* (strain PPE1.0) DNA.

ORGANISM *Pediococcus pentosaceus*
 Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; *Pediococcus*.

REFERENCE 1 (bases 1 to 21838)
AUTHORS Leenhouts, K.J., Bolhuis, A.A., Kok, J.J. and Venema, G.G.
TITLE The sucrose and raffinose operons of *Pediococcus pentosaceus* PPE1.0
JOURNAL Unpublished
COMMENT On May 25, 1994 this sequence version replaced gi:475106.
FEATURES
 source
 1. 21838
 Location/Qualifiers
 /organism="Pediococcus pentosaceus"
 /strain="PPE1.0"
 /db_xref="taxon:1255"
 487..510
 /notes="DR2; putative"
 /rpt_type=direct
 488..561
 /notes="IR1; putative"
 /rpt_type=inverted
 513..561
 /notes="DR1; putative"
 /rpt_type=direct
 2724..2760
 /standard_name="terminator 4"
 /notes="putative"
 complement(2929..3864)
 /gene="rafr"
 complement(2929..3762)
 /gene="rafr"
 /function="regulation raffinose operon"
 /notes="putative"
 /codon_start=1
 /transl_table=11
 /product="regulatory protein"
 /protein_id="AA25562.1"
 /db_xref="GI:475107"

repeat_unit

repeat_unit

repeat_unit

terminator

gene

CDS

/translation="MGEYKTLANKSFESNVLFEGQACLPNTYKGNVRDSYVTHY IQGKGTFAANHPAIVLRAGDIFILPKGTPCYQADNDQPKFYWGFSAGIRISAM LSGSLAQKQVONGHIYADLSLYVLPNSLINDVLGSLIYLFYDLSRRE PADATNIKYKTEQFNLAIVSYLOENTSTGCTIMDLCHYLNLSRYLTLFKTHANTSP QKLLTKRLLEDKQRLSTNNVSQSIANVMGVKDSFTFSKAFKRYSGASPSYRKSG I"

-10_signal

-35_signal

gene

-35_signal

-10_signal

CDS

/function="raffinose transport protein"

/notes="putative"

/codon_start=1

/transl_table=11

/product="permease"

/protein_id="AA25563.1"

/db_xref="GI:475108"

/translation="MQEENYKWWGRLIYGFCAKGNDAFYSLGSLIIFITSHLFD TGNKALDNRWLSIVTLIMVLRIVELFIDPFIQNAIDRKNKSPGHPRPVWVGTVSS IILLFLTLGLGVLAKNAMIYLVFAILYITMDIFYSKDVGVFWSMLPSLTDSRRE KTAFAKLGSTGGGLVGLVMPAVIFFSAKATSCGDNKGFALLICLIASLAWG VGLGTEVSDIRKNQDQIVGWEIFKALAKNDQLLWAAALAYLFYGVGINIGSLEVY YFYIMGKPKSFSLINIFILGLIATSLFPVLKSKFSRKGVFAGCLVFMJGAIPT IAGSNLWLVLAAATMEGFPQMVFLVLMVITDSVEYQKLGLGHRDESALSVRPLID

KREGALSNGVVGQIAIISGMTTGATASSITAGQLHFKLTMFAPFALLMLLIATGIFSK QFLTEKIAEIVAELETRWTRKFDNTQVAKKVVTSLDLATPIAGQVPLAQVNDP TAAQTLGDGFAIKPSDGRILAPFQVATVQVFTTRAVGLVGDNGIIVLHIGLGVK LRGTFISVVEGQHQVQGDLELLEFWDPTIKAGLDDTVIMVTNSTETFTMDWLKRP GQAVKATDNLQLHTRKA"

gene

/gene="agar"

5976..8177

5976..8177

/gene="agar"

/codon_start=1

/transl_table=11

/evidence=experimental

/product="alpha-galactosidase"

/protein_id="AA25564.1"

/db_xref="GI:475109"

/translation="MSLIIVDQANRVFHLHNQTLVIFAVEQCGTSLHLYFGGHVDHY HGLRYPRVDRGFSGNLPGSTDRTFSDTLPKREYSTAGENDYHLPAAIVRHTDGANAL YLVYQGYRIEAGKPLSGLPAAVFEDETEAETLIVLVQVSQVEFDLQTYIYDRPV VTRSVQVCNGDGHAVNLKRVASMQIDFTDQFETITLPGAHANERHPERGSINYGQT EGSRLGTSQHMNPFLALVDHTTTFESGDAYGNLVYSGNHAPELKDOLDLHLAVG INSYNFNOLKAGATQTPTEVLWVYTNKGLNMSQAVHHLIRERVVSSEKNGDERPIV VANWEATPFDNEAKLKPIVDEAKQLGIEMFVLDDGFWGHRDDNSLGDQWVDHRKF PUGLNHFVYVHEQGLKFGIWEPEMISYDSKLIYQOHPDYLMOVGRSPSPSRNQIIL DLGRQAVRNIFDQLLKSQIDYIKWDMNRHLSDIYSVALPPRQGEVHYRYVLG LYELLERLTAYPHILFEGCGGGGRFDAGMAYMPQIWAISONTDAVRLTQYGTSL AYPISLATAHVSVSPNQGTRETSMTSRSAVASGLVGLYELDLTOLSSADKQIVQKQV VOYKOIRPLIQEGEFYRLKSPITSNOANMFVSPOODEALVMVFNLTSAQPSLTTK LVLGNPKLNYQNIATKAIFGGDELMQLGFDVPVYQDYTTKYHFRAVNTEN"

terminator

/standard_name="terminator 3"

/notes="putative"

9403..9426

/notes="DR2; putative"

/rpt_type=direct

10396..10442

/notes="DR1; putative"

/rpt_type=direct

10397..10443

/notes="IR2; putative"

/rpt_type=inverted

10481..10716

/standard_name="terminator 2"

/notes="putative"

complement(10727..11593)

/gene="scrK"

complement(10727..11593)

/function="fructose 6-phosphotransferase"

/notes="putative"

/codon_start=1

/transl_table=11

/product="fructokinase"

/protein_id="AA25565.1"

/db_xref="GI:475110"

/translation="MLLGAIEAGGTGVCATGAENGQVSDRISIPITPTTPTVETMTAVDD YFTTHPDAIGIGSGFPGVGNPHDPKYITTTTPKPGWGFDFLGLKSKQFNILPLYVT TDVNEAAYGESMIGIAKDPNSIYMTIGTGVGAGVISONHIFNGRTHTELGHMRLNRL PGDDFKSNCPYHDICLEGLAGPAGVKRTGKAGKDIPVDDPWPPIITDYIAQACVNLG VAFAPDKIILNGVMNQRLQFLPMIREKFAAYLNGYIEVPLDDIIVPAGLGNNSGIAG GLLLAQAALKNA"

complement(11673..13832)

/gene="agas"

complement(11673..13832)

/gene="agas"

/notes="putative"

/codon_start=1

/transl_table=11

/product="alpha-galactosidase"

/protein_id="AA25566.1"

/db_xref="GI:475111"

/translation="MPVEYDPKTLNLHNDQISYVIQILAHRYPVHRVYFGRVFSKQP YFEPMPGSHAFANDRPFPSVTLSEYSTIGSGDYRQPAYIVKDNANQLLPLILE YTCFSYNDQPIINSRQLPPTVSKHTPTVTLVILHTDAVTKLQMDLNTTIFENQPLILRS

Db 14819 TCGAAACCCAGTGTGTGGCCCAATGTTGCCAAACATGGTGGTTCCTTTATCTTCCAGTCG 14760
Qy 140 catctatgctgaataatgccacggggtgcggagtggttgccagtgctctctcctgcggcgaaga 199
Db 14759 CCTCGATGCCCAACATTTGGGCAAGGTGGCGGCACATTTAGCTATTCTTTGTCGCCACTAAGA 14700
Qy 200 gtgaaaagctcaaggcccttcagagtgcttcaggtgtctccagtgctccagtgcttcttggtattacgg 259
Db 14699 GCCAAAAGCAAAAACCCCTGACTTCTTCAGCCGGGTCTCAGCGTTACTCGGAATACGG 14640
Qy 260 agcctgcgactctcgtgtgaaccttcgctgcgctgcggcttcttcacgtatcggtacggtga 319
Db 14639 AACCTGCTATTTCGGGTCAACCTCAGATCAAAATTTCCATTCTCTTTCAGCGGATG 14580
Qy 320 ccgagcgtactcgttggtcgttgattgacactctttaaatacaagcagtgctcgttggtggcg 379
Db 14579 CCTCAGGAATTCCTTCAGCTTTCTTAGGACTATTCCAGCTTCTGTCGGTGGCGATGGGTC 14520
Qy 380 ctgcaggttctcgtgtgttcttctattgatgctccagatattggtcatgttcttcttggtgt 439
Db 14519 CCGCATCATGATCGATCGATTCGATTGCTTCAAGTCGATCCAGCATTCATGCTCA 14460
Qy 440 gtgcaggtgttaccctcttcctcgcattcgcgctgcggcagcagtgcttcttggtgc 489
Db 14459 GTGCTCATCTCGTTCGTGCTGCGCATTTATCCAACTTTATCTATGCC 14410

RESULT 8
AP001508 296950 bp DNA BCT 10-JAN-2001
LOCUS Bacillus halodurans genomic DNA, section 2/14.
DEFINITION AP001508 BA000004
ACCESSION AP001508.1 GI:10172890
VERSION AP001508.1
KEYWORDS
SOURCE
ORGANISM Bacillus halodurans DNA.
Bacillus halodurans
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
1 (sites)
Takami,H., and Horikoshi,K.
Reidentification of facultatively alkaliphilic Bacillus sp. C-125
to Bacillus halodurans
Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
2 (sites)
Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,
Sasaki,R., Hirama,C., Fuji,F. and Masui,N.
Genetic analysis of the chromosome of alkaliphilic Bacillus
halodurans C-125
Extremophiles 3 (3), 227-233 (1999)
99411980
3 (sites)
Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F.,
Nakamura,Y. and Inoue,A.
An improved physical and genetic map of the genome of alkaliphilic
Bacillus sp. C-125
Extremophiles 3 (1), 21-28 (1999)
99184645
4 (sites)
Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.
Replication origin region of the chromosome of alkaliphilic
Bacillus halodurans C-125
Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
99356711
5 (sites)
Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and
Horikoshi,K.
Sequence analysis of a 32-kb region including the major ribosomal
protein gene clusters from alkaliphilic Bacillus sp. strain C-125
Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
99209008
6 (sites)
Takami,H.

TITLE Genome analysis of facultatively alkaliphilic Bacillus halodurans C-125
JOURNAL (in) Extremophiles in deep-sea environments (Ed.);
HORI KOSHI, K. TSUJII;
: 249-284; Springer-Verlag (1999)
REFERENCE 7 (sites)
AUTHORS Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.
Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125
Extremophiles 3 (1), 29-34 (1999)
99184646
8 (sites)
Takami,H. and Horikoshi,K.
Analysis of the genome of an alkaliphilic Bacillus strain from an
industrial point of view
Extremophiles 4 (2), 99-108 (2000)
20263314
REFERENCE 9 (sites)
AUTHORS Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,
Hirama,C., Fuji,F. and Takami,H.
Characterization and comparative study of the rrr operons of
alkaliphilic Bacillus halodurans C-125
Extremophiles 4 (4), 209-214 (2000)
20426005
10 (sites)
Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,
Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and
Horikoshi,K.
Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis
Nucleic Acids Res. 28 (21), 4317-4331 (2000)
20512382
REFERENCE 11 (bases 1 to 296950)
AUTHORS Takami,H. and Takaki,Y.
Direct Submission
JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and
Technology Center, Deep-sea Microorganisms Research Group; 2-15
Natsushima, Yokosuka, Kanagawa 237-0061, Japan
(E-mail: takami@jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html,
URL: http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html,
Tel: 81-468-67-3895, Fax: 81-468-66-6364)
FEATURES
Location/Qualifiers
source 1..296950
/organism="Bacillus halodurans"
/db_xref="taxon:86665"
/note="alkaliphile"
gene complement(427..615)
/gene="BH0278"
CDS complement(427..615)
/gene="BH0278"
/note="BH0278"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB03997.1"
/db_xref="GI:10172891"
/translation="MRKMGTLVLASIFLVACNADENETESVDQNAEEVEEVEDELE
QDQEEFEDVVEETDDEE"
798..1163
/gene="BH0279"
798..1163
/gene="BH0279"
/note="BH0279"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB03998.1"
/db_xref="GI:10172892"
/translation="MKIFFVAAATFLAASQQAALAEPTVSSQEGVSNADVTTFE
TVDAQSTWVWSRGITNVTDTRKTAYSNVYHRDPSHRSAAQLGNDRVSPGLGPGATT
YASIRGNKLNLPATFVFY"
1334..3475
gene

```
/gene="BH0280"
1334..3475
/gene="BH0280"
/note="BH0280
unknown"
/codon_start=1
/transl_table=11
/protein_id="BAB03999.1"
/db_xref="GI:10172893"
/translation="MHAWSAAYADFAVTPRIDLLGPDIVLSVLEEATHHOVNLIR
DVLLGEGKIAUTHTALLTEKSPFFLEAIPLRHGSRLTSEDTRDGLILSSRRKEDV
EETKGTQLSOTHTVAPLPHQFPNYPYATQVEQAGTVTKEQFSATLADLFNHY
LEDYVELEPSTADDFAVQTEPESHVSTTFDEPKTLAWTLGWTGLMLMFYIMRNGK
TVAYISMHGFSLFKIWRRLIGNACGVVFAVYLFLLPFAFPGWPRYLFVFAAMVTGS
WPPLLLAISFLFVFFSSRLIESIKNTKTLFLIVSQAVKVALVSVIFLLQIA
DRLTMEERNVSTDAWAFKGDYGEFYPHSIGYDQEEENQESGLWQDLSITSELYPVL
NAGALYIDAVNLYEERWMATEYFSEHRSIRVNPYLOAFPLVDENKRIAIOBEE
TKRVLLPYQYDKDKEDIIAYFKEDYATWATMPEYQAESYSHLOEQTFDILWASQEE
VFSNLYVFPNENGLIVDFIIOVLTENNYSYMERDFRGSVGNDFLKLKLDQONTART
YEHYLLRLDLQDDNAKHLVTNEQAQKDISIQRALTDITFLVLTATLAFMTIVQ
SSHLLFAQHKRFLRLRFLCHSLFRAYRNVLFWTLATVVVILGIALYRHSQTYLAIV
VLVLFLETVVTVSVLMRLEQKNKVSVLKGE"
3479..4120
/gene="BH0281"
3479..4120
/gene="BH0281"
/codon_start=1
/transl_table=11
/product="ABC transporter (ATP-binding protein)"
/protein_id="BAB04000.1"
/db_xref="GI:10172894"
/translation="MAAAIVELTDISKFSHSETVLRDFNLITDRGEMIAITGPSGSGK
TTLNLIGLEKEDRTQVRLFGKNKPAVHSKOALQFRKSKIAVLFONFALMDDETYSK
NLDPVLESLGKEDRTQVRLFGKNKPAVHSKOALQFRKSKIAVLFONFALMDDETYSK
DLLEADEPTGSDMANRVILDLICQLNKGMTIVIVTHDPEVADRCDRVITL"
complement(4354..4641)
/gene="BH0282"
complement(4354..4641)
/gene="BH0282"
/note="BH0282
unknown conserved protein in B. subtilis"
/codon_start=1
/transl_table=11
/protein_id="BAB04001.1"
/db_xref="GI:10172895"
/translation="MKIEHVAIWVDLEAKTFYVTVFNCHANSKYHNKDKKPEYFL
TFEGGARLEIMQSGIDKEDQDDRMGWAHIAFSLGSKEAVNQMTGRHNDG"
complement(4740..5537)
/gene="BH0283"
complement(4740..5537)
/gene="BH0283"
/note="BH0283
unknown conserved protein"
/codon_start=1
/transl_table=11
/protein_id="BAB04002.1"
/db_xref="GI:10172896"
/translation="MWTTFYVDAFTNQAFKGNPAACVLPTRDRDDIMQHVASEMN
LSETAFLEPHYQDGLSURWFTPVVTEVDCGHATLASAHILWELDHISAEQPIFYTKSG
ILTASKRSGEILDFPEQPKQENYVNLIDGLQIPLYVGNRRFDYLIETDSEQL
KELNPFSLIEQIDTRGIIIVTSKSTSEYDFISRCFPFPAVGVNEDPVTGSAHCLGYP
WOEKLKNFEFLAYQASKRGKMLKIKLQHRVLELGGQAVTVLRSELL"
6023..7096
/gene="BH0284"
6023..7096
/gene="BH0284"
/codon_start=1
/transl_table=11
/product="nicotinate-nucleotide--dimethylbenzimidazole
phosphoribosyltransferase"
/protein_id="BAB04003.1"
/db_xref="GI:10172897"
/translation="MQKVKEMIGSINELDTKAQQQMEQHLNLTLPKPLGSLGLLESIAI
```

```

KIAGITGHTQPEIDPATIVMVAHDGVAAEGVSAYPSEVTQOMVHNFTIGGAAINVLS
RVSNASVOIVDIGVNGTLOPLINKNRHGTNNMQVSGTSMEREEATAALEVGEVAVQ
AAIONGAKLALGEMGIGNTTASSAMLALEQVPEQIVGFGTGLSHEGKVKVAVIK
RAIEVQPNRAADPIDLVAKVGLIAGILVGAAMRIPVLVDCGFTTAAALTAVR
IAPLCVHYLIASHQSVEDPHVCVKNKLLGLTLPVLNLRILGEGSGTALPFIVRSAIRI
AHEMATFEAGVSGAIEDVGEK"
complement(7227..7766)
/gene="BH0285"
complement(7227..7766)
/gene="BH0285"
/note="BH0285

Query Match      12.5%; Score 138.2; DB 1; Length 296950;
Best Local Similarity 48.0%; Pred. No. 3.5e-27;
Matches 472; Conservative 0; Mismatches 493; Indels 18; Gaps 2;

QY 83 tccgcgcaattgagctggagctgtttaaccagggtggtgatactcttcacgcaaggcat 142
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19855 TCCCGATCATTTTGAACAACCTTTCTGTCAAGAGAGAAGATGTCTATTAAAGCGGTGCAG 19914
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 143 ctatggcctaataatcgccagggtgcgcagtgtttggcagtgatttcttctctggcgaagatg 202
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19915 CCCAGCTGTCTTTTCTCAAGCAGCGCGGCTCTCGGAGTGATGCTGGGAACGAAATA 19974
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 203 aaaagctcaaggcgttcgaggtgttcaggtgtctcgcgtgttcttctgttattacggagc 262
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 19975 AAAAAGCTAAAGACACTTGTGGATCCACTTCCATTACAGCGTTTATTGGAAATTACCGAGC 20034
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 263 ctgcgactcttcggtgtaaccttcgcgtgcgtgcgttccttcacggtacggatccg 322
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20035 CAGCCGTATACGGGGTTTACGTTACCACTGAAGAAGCCGTTTCATTAATGCCGCTTATTAGCG 20094
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 323 cagctatcgtggcgtcttgatgcaccttttaataatacaagcaggtgcgttggcgctg 382
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20095 CGGCTAGGAGGGGCCGATGTTGGGCACATATGGAAGTGTTCGTGTCGCCCCAGCGAC 20154
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 383 caggttctctgggtgtgttcttctattgatgctccagatatggtcatgttcttgggtgtg 442
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20155 CGGGTTTATTACGATTCCGATCTTTTACCCTGAGGATGGAAGAGGGTTTGT-----TG 20208
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 443 cagttgttactcttcttcacgtctcgcgcagcgtatgcttattgaccttaacttggttc 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20209 CTTTTCGTCATCGCGATTATTATTTCATTGTATTGGCAGCTGTCTTCAACGTACATCGTCG 20268
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 503 ggcgcaagcgcagcattgaccagatcaaccgctgctccagctcctgcaggaacacca 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20269 GATTTAAGATCCAGTTGATGACGAGGACACACTTTGATGAGTCAGGATCTCAGAAATG 20328
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 563 aagccgagcagagcaccgcgcagaattttcaaacgattccaccatcatccagggcacctt 622
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20329 AAGTAAAAAGAGAAGATGACAAGAAAGAACCTAGCGCATCTGAAGAGATCAAAAGTCGT 20388
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 623 tgaccggtgaagctattgacctgagcagcgtcagcgtatgccaattgttgcagcgcgaaagc 682
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20389 TAAAGGAGAGGTGGTTCCGCTGACAGAGGTGCAAGACCACGCTTTTTCATCAGGTGCAA 20448
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 683 ttggctcggcgttgccatcgcccccaaacaggggcaggttagtttctcccggtgagtgaa 742
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20449 TGGGGAAAGGTGTTCAGTAGTAAGGCCAAAAGAGGACGCTTTTAGTGGCACCAATAAACCGTA 20508
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 743 agattgtgtgttcctcccatctggccatcttcttcagcttcgcagtcagcaagcgtaggatg 802
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20509 CCGTGACCTCATTATTGAGACAAAGCATGCCATCGGGATTACATC-----TG 20556
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 803 gttccaatgtggatatacttgatgcacattggttttcgacacagtaaacctcaacggcagc 862
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20557 ACAACGGCACAGAAATCTTCATCCATGAGGCATTGACACGGTGCATTAATAAAGGGAGC 20616
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 863 actttaaccgcgtgaagaacagcagggcagatgaagtcaagcagggagcgtctgtgtaaat 922
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 20617 ATTTTACGTCATTTATGACAGGGAGACGAGGTCCGACGTTGGTGTGATGTATTCGTTGAAT 20676
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 923 tcgatattgatgccattaaaggctgcaggttatgaggtaaacacgccgattgtgttttcga 982
```



```

Db 20677 TTGATGTCAGCGCAATTACAGCAGCGGGTATGATGATCATCAGCGCTGTCTGATTACAA 20736
QY 983 attacaagaaacccgacgtgtaaacacttaacggttggtggcgaaattgaagcggggagcca 1042
Db 20737 ATGCAAAACAATTTTCAATGTCACAAACGACAGATAAGCGCGAGGTCAACAAGTGAGGATT 20796
QY 1043 acctgctcaacgtcgcaaaagaa 1065
Db 20797 TGTCTATTACGTCATAAAGTAA 20819

RESULT 9
SXSCRA
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE

JOURNAL
MEDLINE
FEATURES
Source

Location/Qualifiers
1. -2655
/organism="Staphylococcus xylosus"
/strain="DSM 20267"
/isolate="cured strain C2a"
/db_xref="taxon:1288"
/clone_lib="sub-genomic in Staphylococcus xylosus mutant scr1"
/clone="pSE4"
complement(1..121)
/note="ORF3"
/codon_start=1
/transl_table=1
/protein_id="CAA49459.1"
/db_xref="GI:407906"
/db_xref="SPTREMBL:Q56203"
/translation="MKSTTLCIHEFLNTHQRCIHQVILLFSLNKLNDVTLNGY"
complement(132..137)
338..342
350..925
/note="ORF2"
/codon_start=1
/transl_table=1
/protein_id="CAA49460.1"
/db_xref="GI:407907"
/db_xref="SPTREMBL:Q56204"
/translation="MQHIYSEFYNOESKITKLKSVGLIVLHDDVNNNSAEDYIWLQ
QRIDNNELEKWACAYVDKQCYWPHSPYVEWHGCGNAPANDHYIGIERCQSKINGIL
TDQGEENEASYSWTIAALLKKYQLPVARVETVKLHKMFDDTECPARAWSIHLDNAPT
ENIKMLQDYFDIKIKYHYENIKESDMFVVG"
-35_signal
-10_signal
995..1000
1040..1045
1053..2495
/gene="scrA"
1053..2495
/gene="scrA"
CDS
```

```

/note="ORF2"
/codon_start=1
/transl_table=1
/protein_id="EIIscr"
/protein_id="CAA49461.1"
/db_xref="GI:407908"
/db_xref="SWISS-PROT:P51184"
/translation="MNYKSAENILQALGGEDNVEAMTHCATRLRLVLKDEGLVDEKA
LGDMDVVKGTFTGGQYQVILGGTGVNKKVSELEKITGKEASVSVEKGTGKNMPPF
QRFVKMLSDIFVPIPAIVAGGLMGINNLTAPGIFYDNQSLIEVQNGFSGLAEMFN
IFANAPFTLLPILIGFSAKREGGNAYLGAALMILVHPMLMSAYDPTKALEAGKEIP
HWNLFGEINOVGOVOLPMLVATYLTAEKGLKRVPTVLDNLLTPKLAALSTGF
ITFSFVGLTRTLGVLSDGLTWLVEFGAIGLLIFGLLYAPITVITGMHHSFIAETQ
LIADSSGTGGSFIFPIATMNSNAGAAALAAPEIKENKKLKGVAASAQALLGITE
PAMFGVNLKLRYPFIAGVSGSIGSAYIAFFKVKAKALTAGTIPGFSISGQNGMLH
YGIAMIIAFIVAFGVTYALSRYKKRNIEA"
BASE COUNT 909 a 347 c 521 g 878 t
ORIGIN

Query Match 12.2%; Score 135.2; DB 1; Length 2655;
Best Local Similarity 60.1%; Pred. No. 1.7e-26;
Matches 268; Conservative 0; Mismatches 163; Indels 15; Gaps 2;

QY 1 tatgatttcggcggtccagtcgcgctcttcggtctggtctactcaccacatcgctc 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1953 TATGAATTTGGTGGTGTATCGGTGGCTTATATTTGGTTTGTATATGCGCCAAATCGTT 2012
QY 61 atcacgtgtctgcaccagtccttcgccgccaattgagctggagctgtttaacca----- 113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2013 ATTACAGGTATGCATCATAGTTTTATTGCTATTGAAACGCAACTTATCGCTGACAGTTCA 2072
QY 114 -----gggtggatccctatcttcgcaacggcatctctatgctgaatcgcgccaggggtgcg 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2073 TCAACGGGTGGTTTCATTATATATCCCTATTGCAACAATGCTTAATATTGCACAGGAGCG 2132
QY 169 gcatgtttggcagctgttcctcctggcga---agagtaaaagctcaaggccttcgaggt 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2133 GCTGCGCTTGCACGGTCTCTTTATTATAAAGAAATAAATAAAGCGGTAGCCCTCT 2192
QY 226 gcttcagggtctcccgctgttcttggtattacggagctgcgcatcttcggtgtaacctt 285
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2193 GCTGCGAGGTATCAGCTCTATTAGGTATTACAGAACCACTGTTTGTGTGTAAGTTG 2252
QY 286 cgctgcgctggcgttcttcctcatcggtatcggtacgcagctatcggtgcgcttgatt 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2253 AAGCTAAGATATCCATTATATAGGGCAATCGTAGGTTTCAGGTATTGGTTCTGCTATATA 2312
QY 346 gcaactcttaataatcaaggcagttgcttggtggcgctcaggtttcttggtgtgtttct 405
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2313 GCATTCTTTAAGGTTAAAGCTATTGCATTAGGTAGTACTGCAGGTATACAGGATTATATCT 2372
QY 406 attgatgctccagatattggtcatggt 431
||||| ||||| ||||| |||||
Db 2373 ATTAGTGGTCAAAATAATGTTGGTT 2398

RESULT 10
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

LLZ97015 5800 bp DNA BCT 05-APR-1999
Lactococcus lactis cremoris sucrose gene cluster.
297015
297015.1 GI:4581474
enzyme II sucrose protein; fructokinase; sacA gene; sacB gene; sack
gene; sacR gene; sucrose-6-phosphate hydrolase.
Lactococcus lactis.
Lactococcus lactis.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
1 (bases 1 to 5800)
Luesink,E.J., Marugg,J.D., Kuipers,O.P. and de Vos,W.M.
Characterization of the divergent sacB and sacAR operons, involved
in sucrose utilization by Lactococcus lactis
```


Qy	415	ccagatatggtcatgttcttctggtgtgtgagtcagttgattacattcttctcatcgattcgagcga	474
Db	4212	GGGTATAACTTACAAATTTATGATTCGATATTTATTTAGCTTCTTCTTATTTAGCCTTTCTTGT	4271
Qy	475	gcgattgcttgccttcttactgttccgcgcacgcgcagcattgatccagatgcaacc	534
Db	4272	ACCTCAATCTATGTCGCGGATGGAACCTAAATCTATTACGAAGAAAAATAAACAG	4331
Qy	535	gctgctccagtcgctgcaggaacacccaagcagaagcagacccgcgaaatttca	594
Db	4332	AATGC-----AACAACTCAATACCAACCTGAGAAAGTTATTATCATGCCA	4376
Qy	595	aacattccaccatcatccagccaccttgccggtgagctattgcaactgagcagcgc	654
Db	4377	GTTAAAGTGGCAACTCTTGCCTCCGATAAATGATTTGTGATTCCTCTGCTCATGTA	4436
Qy	655	agcattgcatgttggcagcggaagcttggctggcggtggccattgcccacccaag	714
Db	4437	AGTGACCCCTGTTTCTCAAAAGAAATATTGGGAAAGCTATTGCAATCAAGCCTAAATCT	4496
Qy	715	ggcaggttagtttcccggtgagtggaagattgtggtggaattcccatctggccatgct	774
Db	4497	GGAGAACTTTTCTCACCGCGACGCGAGAAATATTATTGCTTATGAAACAGGTCACGCT	4556
Qy	775	ttcgcagttcgcacaaagctgagatggtttccaatgtggaattggtatcttgatcacacattg	834
Db	4557	TATGTTATAAAACAAAAATGGAGGT-----GAAGTCTCTTTTACATATTGGA	4604
Qy	835	ttcacacagtaaaacctcaacgcgcacgcacttcaaccgcgtgaagaagcagggcgatgaa	894
Db	4605	ATAGATACTGTTTCAATGAATGGTAATGATTCATACAAATGTTAAAGTTGGCCAGAAA	4664
Qy	895	gtcaaacagcagggagctgctgtggaattcgatattgattgcaattaaagctgcaggttat	954
Db	4665	GTAAGACGAGGGGATTTTACTAGGATCTTTTGTGATAAGAGAAATCAAGAAGAGTGGATTG	4724
Qy	955	gaggtaacccgcgcgattgttggatttcgaatt	985
Db	4725	GATGATCTGTAATTTGTTATTACAAAT	4755
RESULT	11		
LOCUS	AE004395/c		
DEFINITION	Vibrio cholerae chromosome II, section 52 of 93 of the complete chromosome.		
ACCESSION	AE004395		
VERSION	AE004395.1		
KEYWORDS	GI:9658068		
SOURCE	Vibrio cholerae.		
ORGANISM	Vibrio cholerae		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.		
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C. and White, O.		
TITLE	DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae		
JOURNAL	Nature 406 (6795), 477-483 (2000)		
MEDLINE	20406833		
REFERENCE	2 (bases 1 to 10264)		
AUTHORS	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.		
TITLE	Direct Submission		

JOURNAL	Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .10264
	/organism="Vibrio cholerae"
	/strain="Ni961"
	/serotype="O1"
	/db_xref="taxon:666"
	/chromosome="II"
gene	/note="biotype: El Tor"
	complement(114. .1553)
	/gene="VCA0653"
CDS	complement(114. .1553)
	/gene="VCA0653"
	/note="similar to GB:M76768 SP:P22825 PID:155263;
	identified by sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="PTS system, sucrose-specific IIBC component"
	/protein_id="AAF96554.1"
	/db_xref="GI:9658069"
	/translation="MDIPVIAKQLLESGLGKSNQALAHCASTRRLRLVNDQTQINESA
	IESQGVKQKRVAGQYQIIFSGIVNQVYAEAMKLTGLVEMSTNDVASAAGKONWA
	QRAVGLSDIEFVRIPAIVAGLLGLNFVLTAFGLTEGSLDANGLASMIN
	TFANAFVYLVLLAFASRKFNGNPLGALGLMVMHPDLINGWFGSASVSTVPT
	WNIGFEIEKVYOGSVLPVLSAVILAKIENGLRKIVPSVVDNLLTTHMLAIFTGFL
	TFTVVGPLTRDVGWGLDALNWLVDYSGFVGALGEFIYAPFVITGMHHSIATLQTL
	LADITGCTGTFIPIAAMNSIAQGAALAVGVMKTKLGVAIKVPSTVALLGITEPA
	MVGNLKLRYPIAICGAALASAFITLIVKQAQALGAGLPGIISINPQIGIYYIM
	MAISFAAFALTLLAMREKTQAQVTA"
	1854. .2846
	/gene="VCA0654"
	1854. .2846
gene	/gene="VCA0654"
CDS	/note="similar to GB:L14685 PID:289864; identified by
	sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="sucrose operon repressor ScRr, putative"
	/protein_id="AAF96555.1"
	/db_xref="GI:9658070"
	/translation="MASLHDVARLAGVSKTVSRVINDEYGVKSTKIKVLKAVQCQ
	YMQVAKDLKSQNTLHIVPRVSSHATQVGLTAIFEQAGKHVLLANTHOVHA
	KEIYIOLFNOKRVEGIVFYATHLDOPLVHALQHSAPVPLVQDGSVLYNTPSIYVDD
	HRVFEAGORLVKAGCKQIGFIVQSDDIADVLRSLQGLQALGHQOITLPHARGNF
	SIESGYMAQAYLDIYDLPOLDGLFCATDLRAIKALQELGVHVQGVQQLVGLGNDL
	AVYGNPLSTFNYPADKAGENAAMLLERAGRGQEMSKVVLTFQNIQRETCPV"
	2922. .4568
	/gene="VCA0655"
	2922. .4568
gene	/gene="VCA0655"
CDS	/note="similar to GB:M74035 PID:155269; identified by
	sequence similarity; putative"
	/codon_start=1
	/transl_table=11
	/product="sucrose-6-phosphate dehydrogenase"
	/protein_id="AAF96556.1"
	/db_xref="GI:9658071"
	/translation="MLLDLTLELAGGINNVTRIILAPQGVVIALKHPPLVPHLPDDVS
	LOSVLGEWOLSVQRTAEVSDQALAAIGKAIAROKLETLPYQALDCPYRPLWHISPP
	QGLLNDPNGFIYHOGVEHLFYQHPHFACEDKDYVHLKSLDLVDWQSVALTPTSDW
	FDSHGVSGHAYSHQDQLNLFYTGTRGLVDQRQRTMQCAARMNAGEFEKLGIVRC
	LPEGVTHIRDPKVIYTGKQWMLLAGQTLAHQGLAVIYSDLLHWHFDELKLYDELG
	DYGMWECPPDFELQGEAFVFGPGQIASANPHHTIEHONRIFRATONAQEIALQ
	WPLDEGDFYAPQATQADGRVLCGWMGLPDETQHPSCDQGWTHQTLALRELEWREG
	RIYHPLRELDLTQSEPHTLSDNVTELKTSFALQVTLPGWCELRLMONTQYRVTL
	TLIDENOLLIDRSATQIRQDGTIRELKDSTPVELRLADOSSLEIFINOGEHVMTS
	RIFTPRDASGISLHGASVDKLYYMAPASAFENLVNQV"
	complement(4690. .5661)
	/gene="VCA0656"
	complement(4690. .5661)
gene	/gene="VCA0656"
CDS	/note="similar to GB:M76768 SP:P22824 PID:155263;

identified by sequence similarity; putative"

```
/codon_start=1
/transl_table=11
/product="fructokinase"
/protein_id="AAF96557.1"
/db_xref="GI:9658072"
/translation="MNALVRLSSNHLRSDMSRVWLTDGAVVDLIPDGOQHLKPCPG
GAPANVAVAIARLCGRSAFFGRVNDPFGREMOOTLTDEQVDCOHLFPVHRTSTVV
VLDDEGERSFTFMVKPSADOLFQSLDIPSFONGEWHVCSIALANPSSSTFAIA
QKQVGGYVFDNLRREEVWSEPQELQATVRAGVLADVKFSEELQLFTQSTSEE
GLQATADFQIPLVVTLGKAGALVVTNPSRQIVSGKAVKPIDTTCAGDAFVGLLRL
SVAQDMHNOATLIDAVKWANGCGALATTQKGAMTLPNOALYAE"
```

gene

CDS

```
/gene="VCA0657"
/notes="similar to SP:P13035 GB:M21277 GB:M55989 GB:M96795
PID:I46181; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="aerobic glycerol-3-phosphate dehydrogenase"
/protein_id="AAF96558.1"
/db_xref="GI:9658073"
```

```
/translation="MNARSTRFTQDEVLDLIIVGGINGAGIAADATGRGLKVLGYLDA
KDFASATSSAKLVHGGRLYLHEHFRFLVSEALAEVLLKKAPHIVTPMFRPLPHR
PFLRPMWIRAGLFYDNLGKRTSLPASHKVLKAGSVTKPEMOIGFEYSDCWDDAR
LVILNMQAOEGAEVLNXYCTVEKAERMCDLWHVTLIDQEQTOQERRSHALVNAAGP
WYKQFLNENHVSYPGIRLQSGHIIIVPRIHDEFOAYILQNEDKRIVFIPILODYS
MIGTIDVEYKGRPVAVTDAERDYLISIVNKHFRKREARSIIIAEFSGVRPLCDDDES
NSPQAITRDYLSLDQADAEPLLSIFGKLTYYRKLGEAMKHLAPFPKMKAPWTA
DAPLPGENFYAALKNLQVAAPFETTESVIERWLSYSGSRRTQLLAGVTGEDLGIA
FSGELYQREIDYLCERFARHAQDIFWRRSKLGHLNHTSVVEEVSYLQKQFHAQPL
KATV"
```

gene

CDS

```
/complement(7648..9258)
/gene="VCA0658"
/complement(7648..9258)
/gene="VCA0658"
/notes="similar to GB:X61367 PID:48195; identified by
sequence similarity; putative"
```

```
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="AAF96559.1"
/db_xref="GI:9658074"
/translation="WYTFALIKITLLEQQAEMSAVTPSAQOEVLVGDHDLQVSTDLKG
VTTCNDFCRITAGQADQLLGNHIVPHASMPKRAAFADMMHLLKQGHAWRGIVKNR
TKSGFYWVDAYVTPIYQQQLTGQSVRVKAEKRWELATKAYQALLAEKAGKIQ
FKLHTSLRLALLGALMSPALAHGFOAPEQWQMLASLLPAGVLGLLFRQELVRTPOQL
KOWNEYDSISRLIYSGADAFSVADYHLKMASARITILGRMMSARPLGELANQLHL
TTQEVHQALAAQNSIQAVTQATDAVESAAARVSSHSAHQLIDQVQDHCAETKHSI
NVTHQNLQRLATOAESALTTLKLSDOAQVGLMTETGGIAEOTNLLALNAAIEAAR
AGEQGRGFAVADVEYRALSARTQATQLOLQSIDTLMSTIEAWRGDITASRQDTQCA
QANTLTLOQLQDVECVMSDMLRIVGEVAVAAHQHRELTCVENVQNHISASVATQNSAA
THTVSOLAMWSGKVAEFGLSKQFAQK"
```

gene

CDS

```
/complement(9444..10067)
/gene="VCA0659"
/complement(9444..10067)
/gene="VCA0659"
/notes="similar to PID:2062640; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="protein F-related protein"
/protein_id="AAF96560.1"
/db_xref="GI:9658075"
```

```
/translation="MFEMKFWVLCVTELLTGGSLVTRDMFNNDLLDVPKGFDFVRY
PEWYAPQKNTSVMFNAMTQPSASYEDRLQYLLSNGIHEVYVPGDPMIKLNTVRF
ETGSAKVSMAKQWLDTVARELSDGIDIVLECHTQNTGSEKLNKLAERRANAVKA
ALVQSRVAQNAIYTRGFGENVPACTNSTKNGRACNRREIRFILASN"
/complement(10068..10190)
/gene="VCA0660"
/complement(10068..10190)
/gene="VCA0660"
```

gene

CDS

/note="identified by Glimmer2; putative"

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96561.1"
/db_xref="GI:9658076"
/translation="MNGEHSNHGLCDGKOPICKLLMLLTKDVFNRMPINVKFNN"
```

BASE COUNT 2711 a 2569 c 2371 g 2613 t

ORIGIN

Query Match 11.5%; Score 127; DB 1; Length 10284;
Best Local Similarity 55.9%; Pred. No. 3.5e-24;
Matches 269; Conservative 0; Mismatches 200; Indels 12; Gaps 1;

```
QY 1 tatgattcgccgggtccagtcggcggtctctgtctctgtctctgtctctcaccacaaatgcgc 60
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 656 TATGACTCGGCAGGGTTTGTGCGTGGCGCACATGTTTGGCTTTATCTATGCTCCTCTTTGTG 597
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 61 atcaactggctgcaccagctcctccgccaatgagctggagctgtgttaacca----- 113
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 596 ATCACCGGGAATGACCACATAGCTTTATCGCGATAGAAACTCAGCTACTTGCAGATATCGTG 537
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 114 -----gggtgatccttccttcgcacggcgcctatgctcctaatacgccagggtgcg 168
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 536 ACCACGGGTGGCACCTTTATTTTCCCTATCGCTGCGATGCTAACATTGCGCAGGCGCC 477
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 169 gcattgttggcagtgcttccttcctggcgaaagatgaaaagctcaaggccctgcaggtgcgc 228
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 476 GCGGCTTTAGCTGTGGTGTGATGACGACAGACCAAACTCAAGGGGTGCGGATTCCT 417
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 229 tcaggtgtctccgctattctgtattacagacctgcgacttcggtggaaccttcgc 288
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 416 TCGGGTGTGACCGCGCTGCTGGGGATCAGTGAACCTGCCATGTTTGGCGTGAATTTGAAA 357
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 289 ctgcgctggcgttcttcctcgcgtatcggtaccgcgagctatcggtgcgctttgattgca 348
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 356 TTGCGTTATCCGTTTATTGTGTCGCTGCTGTTGGTGGCGCTTTAGCCAGTCTTTCATCACT 297
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 349 ctcttaataatacaggcagttgcgttggcgctgcaggtttcttgggtgtgtttctattc 408
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 296 TTATTCAATGTAAAGCCCCAACGCTAGGCGCGCTGTGTTGGCAGGGATCAATTTCCGATT 237
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 409 gatgtccagatagtgctcatgttcttggtgtgtgcaattgtaccttcttcacgcattc 468
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 236 AATCCACAGCAAAATGGTTACTACATCATGGGTATGCGGATCTCATTTGTGGCGCGCTTT 177
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 469 g 469
Db 176 G 176
```

RESULT 12

AE006222/c

LOCUS

DEFINITION

AE006222 10085 bp DNA BCT 08-MAR-2001

Pasteurella multocida PM70 section-189 of 204 of the complete

genome

ACCESSION AE006222 AE004439

VERSION AE006222.1 GI:12722266

KEYWORDS

SOURCE

ORGANISM

Pasteurella multocida.

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Pasteurella.

REFERENCE 1 (bases 1 to 10085)

AUTHORS

May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and

Kapur,V.

TITLE Complete genomic sequence of Pasteurella multocida, PM70

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)

MEDLINE 21145866

REFERENCE 2 (bases 1 to 10085)

AUTHORS Zhang,Q. and Kapur,V.

TITLE Direct Submission

JOURNAL	Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA		
FEATURES	Location/Qualifiers		
source	1. .10085 /organism="Pasteurella multocida" /strain="PM70" /db_xref="taxon:747" 344. .625 /gene="PM1841" 344. .625 /gene="PM1841" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03925.1" /db_xref="GI:12722267" /translation="MVNCRIGSLSEILHKQLGHQSIIVNCRIGSLSEKDSRAGRKSYVNC RIGLSLESLCTAFFNSTVYVNCRIGSLSEKLNQLHLFLDYNCRIGSLERIY" 1158. .1400 /gene="PM1842" 1158. .1400 /gene="PM1842" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03926.1" /db_xref="GI:12722268" /translation="MQCASENTDRVNCRIGSLSEIFSQDCLLNPVNCRIGSLSEKQNC IAKCFIVNCRIGSLSEMNVRNATFNVCNCRIGSLEM" /gene="PM1843" complement(2259. .3224) /gene="PM1843" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03927.1" /db_xref="GI:12722269" /translation="MKICLVKTSMDILHSLPALTAQRALPNLQVDVWVVEESFAEI PHWRSVAVQIPIALRWKTPFTALRKQWKYRTLQSEQYDAVIDAQGLLKSALF ATRLARGTRHGYSRHSREPIAACFYDKKYAISYQQHAVERIQLFAQSLNPLPHTQ GDYGIARHPQHAQTMPEVLVFFHCTTRQDKFLPDORWRTLAELKTAQCIQIHPWPSNS QEKORAEIRIAGGLDLVTLPKSTSELAAQIIVNASAVVDVTGLAHLISAALDKTNLT YGATDPKLIGSKGNQHYLVNRESMDKTEADEIFOKLTALLSLNTI" complement(3221. .4267) /gene="rfaf" /note="PM1844" complement(3221. .4267) /gene="rfaf" /codon_start=1 /transl_table=11 /product="Rfaf" /protein_id="AAK03928.1" /db_xref="GI:12722270" /translation="MHILIGPSVGVGMWMSHSYQTLKQHYPLCEIDVMAPNWCKPL LERMPVQAATMPLHGAFALGDRIYRGLKALRHQYDMAIVPLNLSKSAFIPPFKAKIP LRWGKGESRYFLNDLRNSKKDYPMMVQRYIALAFEGKAVPKAEDLPFPVYLTVDP VTQOHTLKPEKQTLALISPRSIGFCFAEFGPAKRWPHYHYAKLAEMLIQQGYAIRL FGSAKDAVGEIRQAALPETMRHFCINLAGQTSLEAIVDLIADCTAIVTNDGLMHTA AAVQRPVLVAGTSPYTPPLSEKAVIIRIEGLGKVRKGDSDGYHQSLIDIQ ELVFDKLTGLLANT" complement(4333. .4743) /gene="PM1845" complement(4333. .4743) /gene="PM1845" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03929.1" /db_xref="GI:12722271" /translation="MKSILTOLLFLTLTALLPTSLFAHIVHQAOFMFSKAGEPSAIFWN IHNTGNEVSLALVQSDIRADIVLHGSTQHGKMIETGINLPAHQMTALKRGGHLIMVF		
gene	gene	gene	gene
CDS	CDS	CDS	CDS

NVEODLNVDGSPFLRLLLFDNGEIMHIEAKIIQY" complement(4859. .6283) /gene="ptsB" /note="PM1846" complement(4859. .6283) /gene="ptsB" /codon_start=1 /transl_table=11 /product="ptsB" /protein_id="AAK03930.1" /db_xref="GI:12722272" /translation="MNFQIAQQVIDKLGGKDNISAAAHCAATRLRIVVKNENLIDKKG IENTEGYKQFAVAGYOIIFGSGTVNKVYAALSKLIGIGDMTSEVAAAGTEKQGLL ORLVKGLADIEVPIIPAIVAGLLMG IHSMLTAGFFVEKNVVDLYPAIADLVDFIN TIANAPFVFLVILGFSATRGGNPFLGAALGMILLVHPALSDGNVALTAKNGIOY WHIFGLEIRVGYQGTIVPLVASWVLTATLERNLKKVYPSFDNLITPLFLAFTIGLL AFTVIGPIREGAGSLISLGLTWLXDTLFGVGAIFGTLYAPIVIVGMHQTFIATVEQL LAEYVGRYTFPIITAMNSNIAOGAAGAAVYMKDAKRVGIAVPSGSIALLGITPEA MFGVNLRYRYPFISAMIGAGISSAVIALFNVKATALGAAGLPGIPSIKPDLSLAMYCVG MLISASTAFTLTVLGKRAQLKAE" complement(6512. .7531) /gene="scrf" /note="PM1847" complement(6512. .7531) /gene="scrf" /codon_start=1 /transl_table=11 /product="Scrf" /protein_id="AAK03931.1" /db_xref="GI:12722273" /translation="MSITNKKRITLSLIAKCCGLSTTVSMILNNRGEEFKIKAETCQ RVKATAEQGYRANVYAKSLKORSNVIGFVLPDLTNYGFACTAKILKOCRENGLOL IIACSDNNPQEKALIERLLDRQIDFLITAPTHQDPSYQKIGQHTPIQLDRYVPL ALNYVISDNSEKIALICRLCDYQLTETEFYMGQRTLSPSESRLAGLQGLKQANLN LKEMILHRTYQPSGYQMLAEAKGLRQLPAIFTASTYTLLEGVLYLTQKKMDLL LNROLHLATFDHDLNALPHIHSIKQNHPLAHQIFQHIWRYFAGKSPNQKIDCD VIYRP" complement(7557. .9056) /gene="scrf" /note="PM1848" complement(7557. .9056) /gene="scrf" /codon_start=1 /transl_table=11 /product="Scrf" /protein_id="AAK03932.1" /db_xref="GI:12722274" /translation="MPFSLSKVRSPFCHFWTSPMYIFNQGKYKSLHAQOVGELAQIRO HVLRDTPRPTVHLAPENGLLNDPNGLVFDGKGYHI FYOWPPYAAIHGMKKWKHPMTY DFHFQVADPLIPDELFEHSGCYSGALLWQDOIYAVFYGTNRNAENORIPHONIAIF KKDGTLLAKRCLLNOAPTGYTEHVDRPKPFIIEBGRIFILGAQREDLSGALITYEMT DLDATPQLIGBEFLKLPDFDNQVFMWECPDLLKQEDLFLMSPOGKAQTYCFQNHYY ATYALGOYANLHFSQAHEEDYDFDFAQOSIQGSEGIYAYWGLPDLTYPTDEYQW HMSLSLPROFSIQOGLIYORPIPOVYAQCLPAHYWVEGKQOI PHLDRSYVAFDAENQ PFTREFTNQSQOGLILRYQNGILTLDRYTTOTALMQQFGEORHCILPOLNHHVEIFF DRSYVEFLNAGEKWMTSRFEIADRENALDSSRLCVSVLAEKPIILTP" complement(9037. .9963) /gene="PM1849" complement(9037. .9963) /gene="PM1849" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK03933.1" /db_xref="GI:12722275" /translation="MGKIWLVLGDVVDLIPGENTYLRACAGAPANAVAGIARLGATT GFIGRVGDDPLGREFMLANKLOENVCTQYMHLDPKQRTSTVLDLKEGERSFTFWNPS ADQFLELSDLPFPNPDHCCSIALINEPSRSSTFAAIHRIKAGGGFFSDPNLRES LWNSPAEMKTVVNVKALADYLVKFSSEELLEFLDSDATAKLTQHPDKLLIIVTL KMGATYTYLQGSQVLSVKKALIPDVTGADGAFVSGLLFGISQVKNQNPNTALVEII KQANACALATTAKGMSALPNKQOLDAFLAE"			
gene	gene	gene	gene
CDS	CDS	CDS	CDS

BASE COUNT 2923 a 1988 c 2212 g 2962 t
ORIGIN


```

DTEPGSLAAMIDLTKKAEPGGLPFAISDSEAMKERRIKTKTKTCTCTCGVCSFEVW
TKDRILKQVSHDSPANKIVTCVKGFESWGHISNDORLTKPLVRKNGEFHEVEWDEA
LNVIAADNTLSIKEKQPDALSPFISSKATNEHSYLMOKLAROVIGTNVNDVCSRYCOA
PATKGLFRTVCHGDSGSTEDLEKAAMSVLIGTNTAEAPVIAKRAQLFGQKIH
VFDIRKHAERADRFQYKPGTQDLAWLSAVTKYLIIDHDLKRAFIDEWDDFDIYIK
SLETFMAAEATGIPESLTKFAEACAKASVVICWAMGITQODIGSDSSTAISNL
LLVTGNRRPPTGAYPLRGHNHNVQCSMDMPKTIQYQSTEAADIIRAKFEKEYGVK
LNPAGKNHEVGEIHGDEHSLVLYGDTGTIVDSNINVFQAAPEKIDPMVQDEFL
TFTATYADVLPASPSEKDGFTFTERRIORLYOALEPLGDSKPDWKIFQAIANRLG
EDWNAKHSEINDEVARLITPLAGVSYDRLEGFNSLQWPVQDGDDEPILYLEGFNFD
NGKALFLPSPNFQDEYIDIHVNNGRLLHFPHEGNTYOTPMIKYKVPRAFVEIS
PELADRGHEGAAYKLISSETGEAVLVHVDTRVKGEIYIPLNDAMENGDLGAINL
ITNSVDQYDTPTSPYKRTLSCELEVTIKRGKSLPNPNFRVKKRHPQYSVQVQKKWER
SDYVFPNGVDR"
/translation="7804..8751)
/gene="SAV2310"
/complement(7804..8751)
/translation="SAV2310"
/translation="similar to lyt divergon expression attenuator LytR"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58472.1"
/db_xref="GI:14248084"
/translation="MKRSKSKMSLPKFIKFLVFGILVLAIVVYVAAKIFITGNK
IHNPLDRHSELKRAEIVRGDTTEKIAHAYAGGPNMAKVKSLKLMNVPIIDHYATIDM
DGLHNMIDSIGDVVDVNDFTVDGVRTKGOOTHVNGDOALKFIRSRKEGAGDGF
RQORQIVLEAMANKIASPSSITHENSMLNEIQNVKTDLILGLDNLIRSNYKANDT
INKHLSQGGGQSDGLYFIPSEQSAESTKLLKNDLE"
/complement(8884..9681)
/gene="SAV2311"
/complement(8884..9681)
/translation="similar to suppressor protein subh"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58473.1"
/db_xref="GI:14248085"
/translation="MDKTLQIDKILCSWLKQIDNVIPOLIMETTTETKRHRFLVLT
NVDKIQOQOFLATHPEHQLLAEKSNEMITNEIHNHLMIDPIDGTANLVKQOED
YCIILAYEYKPMLSYVYDYPHKKLYKAIIRGEAFKMEEPSPSLKLEDAIISFN
AQVMNLDTVQDLDFDASFSYRLVAGCLDSMRVAKQFGAHINTNPKWDIAAQFLFAE
LLNLKMTLDGKAIDHLKGAPEIISNKACHETVLAILNANGYQYR"
/10066..10758
/gene="SAV2312"
/10066..10758
/gene="SAV2312"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB58474.1"
/db_xref="GI:14248086"
/translation="MKAERONLIITAIQONKMTALELAKYCNVSKRTILRDIIDLE
NOGVKIYAHYCKNGCYOIQOAKSIALNSETOLSALFLVLNESOSYSTLPVKSEINA
IILKQCLSPOTRLRLKRMDEYIKEDTOHTMLPMLFSDILLYCTERNVLMVDRVD
DNKAEVIFGLCKLGHGHWAVIYDIAQDKTAELEIENIDISYSFKTIQTRDISI
DNYHFLNPIDS"
/complement(10783..11520)
/gene="SAV2313"
/complement(10783..11520)
/translation="similar to protein of pX02-46"
/gene="SAV2313"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB58475.1"
/db_xref="GI:14248087"
/translation="MENEKKTYTFKDIARWDLIIPIIGVIFALTIAAYIGPILFGY
YEIDNSLIFVATIQALAYTIGIGAFYFLHKLMPSRSLRAGFAYIKKWLRLSIYII
ALILITYEYKQFLQYSETANELEKNMFVPAFLPVAFLLIIVGPIVEEIV
FRHILIGELGRKFNFIAMSVSVFLFAFIHVTDAKSPFEFGPYLILSIILVFTYKLSG
```

```

RNLGSTIALHIANNEVSVFVISIYA"
11569..11898
/gene="SAV2314"
11569..11898
/gene="SAV2314"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:14248088"
/translation="MINETKMYRQOITNLIQIIDIRIKYNTLMDNIOHTHIYKFNQIV
TNELNKFQFFETSYIINENQVSYEIIILLNKRPLRQVDYGNKIQLYLNPHYFELSAL
FAIKFAH"
12343..13215
/gene="SAV2315"
```

```

Query Match      10.9%; Score 121; DB 1; Length 341350;
Best Local Similarity 55.9%; Pred. No. 2.1e-22;
Matches 279; Conservative 0; Mismatches 205; Indels 15; Gaps 2;

QY 1 tatgatttcggcggtccagtcggcggtctgctctcggtctggtctactcaccacgtc 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76239 TAIGAATTTGGTGGCAATTTGGTGGAATTAATATTCGGATTATTGTATGTCGCAATGTT 76180

QY 61 atcactggtctgcaccagtcctcccgccaattgagctggagctgtt-----t 108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76179 ATTACAGGTATGCATCATAGCTTTATAGCTGTAGAAACGACATTAATTGCAGATGCCACT 76120

QY 109 aaccagggtggatccttcattcttcgcaacggcatctatggcctaataatcgccccagggtg 168
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76119 AAAACGGTGGATCATTTATATTCGGATTTCGACAAATGCTCTAATGTTCACAAAGTGGT 76060

QY 169 gcatgtttgagctgtctctcctggcgga---agagtgaaaaagctcaaggcccttgcaggt 225
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76059 GCAGCAATTCGACCGTCTCTTATTATTAACAAAAATAAGAGTTAAAGGTGGCATCT 76000

QY 226 gcttcaggtgtctccgctgtcttggtattacggagcgtcgatctctcgggtgaaacctt 285
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75999 GCGCAGGTATTTTCAGCATTTACTTGTGTTATACAGAACCGGCTATGTTGGTGTAACTTA 75940

QY 286 cgcctgcgtgcgctgtcttcctcgtatcggtaccgcagcgtatcggtggtggttgcatt 345
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75939 AAACTAAGATATCCATTTATTGGCGCTATCGTTGGATCAGGTATTGTTTCAGCATATATT 75880

QY 346 gcactcttaataatcaagcagcttcgcttggcgctgcaggttcttgggtgtgtttct 405
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75879 GCTTTCTTCAAGTTAAAGCAATCGCATTAGGAACCTGCGGATGCCAGGATTTATTTC 75820

QY 406 attgatgtccagatatggtcatgtcttctggtgtgtgcagttgttaccttcttcacgca 465
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75819 ATCAATCCAGTACATGCAGGATGGTTACACTACTCTTTGTTGGTATGACAATATCATTCATC 75760

QY 466 ttcggcgagcgattgctt 484
||||| ||| ||| |||
Db 75759 ATTGCTATAACAGTACTT 75741
```

Search completed: March 22, 2002, 08:55:40
Job time: 8049 sec

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	666	36.6	650	2	Q9S6S6	Q9S6S6	lactococcus
2	562.5	30.9	630	2	Q9KG19	Q9KG19	bacillus ha
3	550.5	30.2	636	2	Q9KF90	Q9KF90	bacillus ha
4	539.5	29.6	674	2	Q99XQ4	Q99XQ4	streptococc
5	532.5	29.2	627	2	Q9LHG6	Q9LHG6	clostridium
6	517	28.4	644	2	Q9KJ80	Q9KJ80	streptococc
7	494	27.1	620	2	Q99Y91	Q99Y91	streptococc
8	480.5	26.4	616	2	Q46129	Q46129	clostridium
9	479	26.3	577	2	Q9L461	Q9L461	lactobacill
10	468	25.7	621	2	Q84808	Q84808	klebsiella
11	448	24.6	636	2	Q9CFK9	Q9CFK9	lactococcus
12	445.5	24.5	620	2	Q9A0X4	Q9A0X4	streptococc
13	433.5	23.8	618	2	Q9X565	Q9X565	enterococc
14	418.5	23.0	480	2	Q99RQ0	Q99RQ0	staphylococ
15	416	22.8	479	2	Q9KLT8	Q9KLT8	vibrio chol
16	411	22.6	661	2	Q68468	Q68468	corynebacte
17	398.5	21.9	640	2	Q9RLJ2	Q9RLJ2	listeria mo
18	390.5	21.4	683	2	Q46072	Q46072	corynebacte
19	350.5	19.2	470	2	Q9KAS1	Q9KAS1	bacillus ha

[illegible]

RESULT	4	
Q99XQ4		
ID	Q99XQ4	PRELIMINARY; PRT; 674 AA.
AC	Q99XQ4;	
DT	01-JUN-2001	(TrEMBLrel. 17, Created)
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)
DE	PUTATIVE PTS SYSTEM ENZYME II.	
OS	SPY2097.	
GN	Streptococcus pyogenes.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI_TaxID=1314;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=SF370;	
RX	MEDLINE=21192684; PubMed=11296296;	
RA	Ferratti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.	
RA	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,	
RA	Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,	
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;	
RT	"Complete genome sequence of an M1 strain of Streptococcus pyogenes."	
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).	
RL	EMBL; AB006630; AAK34748.1; -.	
SK	Complete proteome.	
QW	SEQUENCE 674 AA; 72490 MW; 4BAA26840E784878 CRC64;	

Query Match	29.6%	Score	539.5;	DB 2;	Length	674;
Best Local Similarity	36.6%	Pred. No.	2.5e-25;			
Matches 123;	Conservative	67;	Mismatches	123;	Indels	23;
Gaps						8;

Qy	5	GPVG---GLLFGVLVSPVITGLUHQSPFPIELFLNQGG---SFIFATASMANIAQGAAC	58
Db	322	GPVKWLFGAIFGALVAPLVITGLHMTNAIDTQIADTATRTGLWPMIALSNIAGGSV	381
Qy	59	LAVFFFLAKSEKLG-LAGASGVSAVLGITEPALFGVNLRLRWPFFIGTGAAGGALIAL	117
Db	382	FAYLLMNRHIEREAETSLPAAISAYLGVTEPALFGVNVKYVPFVAGMTGSGIAGLLSTT	441
Qy	118	FNTKVALGAAGPLGVVSDADPMWNFLVC---AVTFFIAFGAAIACLYLVLRNGSI	173
Db	442	FNQANSJGVGLPGFMAINVKTMIPFFTCMAVAIVVPMFLFFFRKSHMTKTEDEAKL	501
Qy	174	DPDATAAPVAGTTKAEAPAEFSDNSTI-IQAPLITGEAIALSSVSDAMFASGKLGV	232
Db	502	-PETPVSAPVAT-----APHKTMQGTVITLTSPLTGEVKALSEAVDPVPAQVWGGA	554
Qy	233	AIPTKGLVSPVSGKIVVAPFSPGHAFVTRKAEDGSNNVDILMHIGFDTVNLNGTHFNPL	292
Db	555	LLQPTGEVLVAPCAEVSVLFTPKHAICLVLT---EGLELLMHIGMDTVNLDGGFEAL	610
Qy	293	KKOGDEWKAAGELLCEFDIDAIRKAAGEVETTPIVWSN	328
Db	611	VKOGDQVKAGQTLQIOPDIAAISAGVATETPLVVTN	646

RESULT 5
Q9L8G6
ID Q9L8G6 PRELIMINARY; PRT; 627 AA.

AC	Q9L8G6;	
AD	01-OCT-2000 (TrEMBLrel. 15, Created)	
AE	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
AF	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
AG	SCRA.	
AH	SCRA.	
AI	Clostridium acetobutylicum.	
AJ	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;	
AK	Clostridium.	
AL	NCBI_TaxID=1488;	
AM	[1]	
AN	SEQUENCE FROM N.A.	
AO	STRAIN=ATCC 824;	
AP	MEDLINE=20391269; PubMed=10937490;	
AQ	Tangney M., Mitchell W.J.;	
AR	"Analysis of a catabolic operon for sucrose transport and metabolism	
AS	in Clostridium acetobutylicum ATCC 824.";	
AT	J. Mol. Microbiol. Biotechnol. 2:71-80(2000).	
AV	EMBL; AF205034; AAF35839.1; -	
AW	InterPro; IPR001127; PTS_EIIA.	
AX	InterPro; IPR001996; PTS_EIIB.	
AY	InterPro; IPR003352; PTS_EIIC.	
AZ	InterPro; IPR001254; Trypsin.	
BA	Pfam; PF00358; PTS_EIIA_1; 1.	
BB	Pfam; PF00367; PTS_EIIB; 1.	
BC	Pfam; PF02378; PTS_EIIC; 1.	
BD	ProDom; PD002243; PTS_EIIA; 1.	
BE	PROSITE; PS00371; PTS_EIIA_1; UNKNOWN_1.	
BF	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.	
BG	SEQUENCE 627 AA; 67596 MW; B835AB0238FCA436 CRC64;	
BH		

Query Match	29.2%	Score 532.5;	DB 2;	Length 627;
Best Local Similarity	34.7%;	Pred. No. 6e-25;		
Matches 118:	Conservative	60;	Mismatches 141;	Indels 21;
				Gaps 4;

Qy	1	YDFGPPVGLLFGVLVSPVIVITGLHOSFPPIELFNOGG---	SFIATASMANIAGAA	57
Db	286	YNTGPPSGFLGGLYSLIVITGTHHSHFAIEAGLANPAIHKNFLLP	PIWSMANVAQGA	345
Qy	58	CLAVFFLAKSEKLGLAGAGSVNAVLTGTEPAIFGVNLRWLPFF	TGICGTAAGGALIAL	117
Db	346	ALAVYFTRDKMKSTAAPASFCLLTGTEPAIFGVNLRKTPF	IAGALGAGIGGYIVF	405
Qy	118	FNIKAVALGAAGFLGVVSDAPDMVMFLVCVVTFFTAFGAIA	TGLYLRNGSIDPDA	177
Db	406	TKVAMTAGVGTGPIGIAIVKQGSFLNVIAMI---	LAFGGAFTIAMVLGKEITEEDL	461
Qy	178	TAAPVPAGTTKAAEAPAEFSNDSTIIOALPTGEAIALSSVS	DAMPAGSLGSGVAIVPT	237
Db	462	NKETVNDIKVEEVES-----	VVSPNCKVLLKNVPDKTFAEGLIGDVGVDPE	511
Qy	238	KGQLVSPVSKIIVAPSPGHAFVTRKAEODGNSVDILMHIG	FDTVNLNGHFNPLRKQGD	297
Db	512	DGEVSPIDGTVVHVFTKHAIAWKSK---	NGVEMLIHIGIDTVKMEGNKFSFINDGE	567
Qy	298	EVKAGELLCEFDIDAKAAGYEVTTPIVWSNYKKTGPVNT	337	
Db	568	EVKKGDKLIQFDLVKEKAVSPIVITIVNTHEDMGFVNS	607	

RESULT	6	
Q9KJ80		
ID	Q9KJ80	PRELIMINARY; PRT; 644 AA.
AC	Q9KJ80;	
DT	01-OCT-2000 (TEMBLrel. 15, Created)	
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)	
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)	
DE	BETA-GLUCOSIDE-SPECIFIC EII PERMEASE.	
GN	BGLP.	
OS	Streptococcus mutans.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC	Streptococcus.	

```

OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NG8;
RX MEDLINE=20340959; PubMed=10878120;
RA Cote C.K., Cvitkovitch D., Bleiweis A.S., Honeyman A.L.;
RT "A novel beta-glucoside-specific PTS locus from Streptococcus mutans
RL that is not inhibited by glucose.";
RL Microbiology 146:1555-1563(2000).
DR EMBL: AF206272; AAF89975.1;
DR InterPro: IPR001127; PTS_EI1A;
DR InterPro: IPR001996; PTS_EI1B;
DR InterPro: IPR003352; PTS_EI1C;
DR Pfam: PF00358; PTS_EI1A.1; 1.
DR Pfam: PF02378; PTS_EI1B; 1.
DR Pfam: PF02378; PTS_EI1C; 1.
DR ProDom: PD002243; PTS_EI1A; 1.
DR PROSITE: PS00371; PTS_EI1A.1; UNKNOWN_1.
SQ SEQUENCE 644 AA; 69282 MW; 723B7FBBDD2794EB CRC64;

Query Match 28.4%; Score 517; DB 2; Length 644;
Best Local Similarity 35.0%; Pred. No. 5.5e-24;
Matches 126; Conservative 65; Mismatches 149; Indels 20; Gaps 8;

QY 1 YDFGGPVGLGLGVSPVITGLHQSFPPIE--LELFNOGGSFIFATASMANIAQGAACL 59
Db 299 YGFNPVYIGVILGAMQVLMFGLHGLVPLALELQKGPVILVATIAIC-FAQAGSLL 357

QY 60 AVFFLAKSEKLKGLAGAGSVAVLGTETPAIFGVNLRWRPFPGTIGTAAIGALIALFN 119
Db 358 NIMRNTNENKVRQLSPAFISALFGVTEPAIYGITLPMRVPTMTCVSGAISGAYLALFN 417

QY 120 IKAVALGAAGFLGVSDIDAPD-----MVMELVCAVVFETFAAGAAIAYGL-YLVRNGSID 174
Db 418 VKMQVGGMGLFAIPSEIDPKNSMLIHFLIAIAMPFVLGVTQFIKIPYLGEPTSTD 477

QY 175 PDATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEAIALSSVDSAMFASGLSGSVAI 234
Db 478 SDVDDKEEPVKELK-----EIKQE--IISPLCKVKLENVPDEVFASGAMKGIAI 528

QY 235 VPTKGLVSPVSKIVVAPPSGHAFVTRKAEDGSNVYDILMHIGFTVNLNTHFNPLKK 294
Db 529 DPDDGIVAPTKGEVLTFTKHAVALRT--ENGA--EILIHGMDTVSLAGKGFKSFE 584

QY 295 QGDEVKAGELLCFEDAIKAAGYEYVTPPIVVSNNYKKTGPVNTYGLGETEAGANLLNVAK 354
Db 585 VGDHVEAGOTLLEFDVNAIKAGLPVITPVIVTNSQDFEDVLTQERTVEAGNYLLTAVK 644

RESULT 7
ID Q99Y91 PRELIMINARY; PRT; 620 AA.
AC Q99Y91;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE PUTATIVE SUCROSE-SPECIFIC PTS PERMEASE, ENZYME II.
GN SCRA OR SPV1815.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savić G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
```

```

DR EMBL: AE006608; AAK34540.1; -.
KW Complete proteome.
SQ SEQUENCE 620 AA; 65725 MW; 64F5FE83524DC4DC CRC64;

Query Match 27.1%; Score 494; DB 2; Length 620;
Best Local Similarity 36.8%; Pred. No. 1.4e-22;
Matches 126; Conservative 47; Mismatches 149; Indels 20; Gaps 5;

QY 3 FGGPVGGLLGLVYSPVITGLHQSFPPIEELF-NQGGSFIFATASMANIAQGAACLAV 61
Db 295 FG--IAGLVGGIQQQLIVVTGIHHIFNLEAQLIANTKDPFNAYLTATAAAGATLAV 352

QY 62 FFLAKSEKLKGLAGAGSVAVLGTETPAIFGVNLRWRPFPGTIGTAAIGALIALFNK 121
Db 353 AVTKSTKLGAFPTLSALLGITEPAIFGVNLRPKVFSGLIGALGCGWAGLFGIA 412

QY 122 AVALGAAGFLGVVSDIDAPDMVMELVCAVVFETFAAGAAIAYGLYLVRNGSIDPDATAAP 181
Db 413 GTGFGITVLPGLTLLYLNGLQLQVLTMLVGLGVAFAIAYTWGY-----QDRETL 462

QY 182 VPAGTTKAEAPAEFSDNSTIIQAPLTGEAIALSSVDSAMFASGLSGVAIVPTKGQL 241
Db 463 LPAVEVDQADQAPALAE--TLYSPLNGTVVDSAVDPVFSGGAMGQGLAIKPEDNTL 519

QY 242 VSPVSGKIVVAPPSGHAFVTRKAEDGSNVYDILMHIGFTVNLNTHFNPLKKQGVDEKA 301
Db 520 YSPVSKVEIVFTGHAYIAITS-----SQGAEVLLHIGIDTESMAGDGFESLVAGQAVK 575

QY 302 GELLCEFDIDAKAAGYEYVTPPIVVSNNYKKTGPVNTYGLGEI 343
Db 576 GDLGHGFDPSKIAEAGLDDTMMIVSNIADYQSDVILAQGHV 617

RESULT 8
ID Q46129 PRELIMINARY; PRT; 616 AA.
AC Q46129;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE PTS-DEPENDENT ENZYME II.
GN ABGF.
OS Clostridium longisporum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1523;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6405;
RX MEDLINE=98151780; PubMed=9491080;
RA Brown G.D., Thomson J.A.;
RT "Isolation and characterisation of an aryl-beta-D-glucoside uptake and
RL utilisation system (abg) from the gram-positive ruminal Clostridium
RN species C. longisporum.";
RN Mol. Gen. Genet. 257:213-218(1998).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B6405;
RA Brown G.D., Thomson J.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: L49336; AAC05713.1; -.
DR HSP: P20166; IGPR.
DR InterPro: IPR001127; PTS_EI1A.
DR InterPro: IPR001996; PTS_EI1B.
DR InterPro: IPR003352; PTS_EI1C.
DR Pfam: PF00358; PTS_EI1A.1; 1.
DR Pfam: PF00367; PTS_EI1A.1; 1.
DR Pfam: PF02378; PTS_EI1B; 1.
DR ProDom: PD001476; PTS_EI1B; 1.
DR ProDom: PD002243; PTS_EI1A; 1.
DR PROSITE: PS00371; PTS_EI1A.1; UNKNOWN_1.
DR PROSITE: PS01035; PTS_EI1B_CYS; 1.
```

SQ SEQUENCE 616 AA; 65890 MW; 7D9CDB1B7BE9283E CRC64;

Query Match 26.4%; Score 480.5; DB 2; Length 616;
Best Local Similarity 31.9%; Pred. No. 9e-22;
Matches 115; Conservative 62; Mismatches 151; Indels 33; Gaps 6;

QY 1 YDFGGPVG-LGLFGLVSPVITGLHOSPPPIEL-ELFNQGGSFIFATASMANIAOGAAC 59
DB 278 YNFNTIAGLFIGGQVQVFMFGLHWGLVPIAMNNAVIGYDPLVATVAVCFAGTGVVM 337
QY 60 AVFFLAKSEKLKLAGAGSVAVLITEPAIFGVNLRWRPFIFIGTAAIGGALIALFN 119
DB 338 ATLAKTKOKKLKLCIPALISGFFGVTEPAIVGITLPRKKPFIILSCIAAGYTGGIIGFFE 397
QY 120 IKAVLAGAAGFLGVVSDAPD-----MVFVLCVAVTFEAFGAIAVGLYLVRNG 171
DB 398 SKGYSMGGLGIFALPSYNPEGIDRGFGVMVAVIGVVGIL-----MFVTKLND 449
QY 172 STDPTAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGCEATLSSVDAMFASCKLGSG 231
DB 450 EEEVKTTSKEESLVKQEE-----IVSPIQGVVTLAEVKDEAFSSGALGKG 497
QY 232 VAIVPTKGQVSPVSGKIWVAFPPSGHAFVARTKAEDGNSVDILMHIGFDTVNLNTHFNP 291
DB 498 VAINPIEGKVYAPADGTLTLFPLSHALGIT--ENGA--ELIIHVGMDTVQLECKHETA 553
QY 292 LKQDEVKAGELCEFDIDAIAKAGYEVYTPDIVVSNYKKTPGVNVTYGLGETEAGANLIN 351
DB 554 KVKQDGKIKKQGLLETFDKAEAKAGYSTIPIVLITNSDYLDVDTDKRKVDVNSSELLT 613
QY 352 V 352
DB 614 V 614

RESULT 9
Q9L461 PRELIMINARY; PRT; 577 AA.
ID Q9L461
AC Q9L461
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PERMEASE.
GN BGIP.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B21;
RX MEDLINE=20263773; PubMed=10802183;
RA Marasco R., Salatiello I., De Felice M., Sacco M.;
RT "A physical and functional analysis of the newly identified bgIPCT
operon of Lactobacillus plantarum."
RL FEMS Microbiol. Lett. 186:269-273(2000).
DR EMBL; AJ250202; CAB71150.1;
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA.1;
DR Pfam; PF00367; PTS_EIIB.1;
DR Pfam; PF02378; PTS_EIIC.1;
DR ProDom; PD002243; PTS_EIIA.1;
DR PROSITE; PS00371; PTS_EIIA.1;
DR PROSITE; PS00371; PTS_EIIB.CYS.1;
SQ SEQUENCE 577 AA; 60877 MW; E8C877050103B5FC CRC64;

Query Match 26.3%; Score 479; DB 2; Length 577;
Best Local Similarity 34.1%; Pred. No. 1e-21;
Matches 118; Conservative 58; Mismatches 88; Indels 82; Gaps 11;

QY 1 YDFGGPVG-LGLFGLVSPVITGLHOSPPPIEL-ELFNQGGSFIFATASMANIAOGAAC 58
DB 285 YNF-MPIGAGVIGAFWQVFIQFWHTFVPLMMNIAKMGVDPLPLISAVALSQAGAA 343
QY 59 LAVFFLAKSEKLKLAGAGSVAVLITEPAIFGVNLRWRPFIFIGTAAIGGALIALF 118
DB 344 LAVFLKSRDQMKALAGSSFTALFGITEPTIYGVTLKLRPFYCAVVGALGGAII--- 400
QY 119 NIKAVALGAAG-----FLG-----VVSIDAPDMVFLVCAVVTFFTA 156
DB 401 -----GAAGTHASSTFLPSLLAVPTFLGHGFMGEVIGL-----IVAFGLGAILTYFFGF 449
QY 157 GAAIAYGLYLVRNGSIDPDATAAPVPAGTTKAEAEAPAEFS-----NDSTIIQAPLTGE 211
DB 450 A-----RQQADYVSVSDQSDNTVM-APVEGT 475
QY 212 AIALSSVDAMFASCKLGSGVAIVPTKGQVSPVSGKIWVAFPSGHAFVARTKAEDGSMV 271
DB 476 IIPLTSHDEVFASAMGKGLAIVPNKGTVPADGTITAVVPTGHAIGITA---NSGA 531
QY 272 DILMHIGFDTVNLNTHFNPLKQDEVKAGELCEFDIDAIAKAG 317
DB 532 EELIHIGINTVOLNGOYFETVMYKQNVKRGDLTKFDVDKIKTAG 577
RESULT 10
Q48408 PRELIMINARY; PRT; 621 AA.
ID Q48408
AC Q48408
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CELLOBIOSE-SPECIFIC PTS PERMEASE.
GN CASHA.
OS Klebsiella oxytoca.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2;
RX MEDLINE=97176370; PubMed=9023916;
RA Lai X., Davis F.C., Hespell R.B., Ingram L.O.;
RT "Cloning of cellobiose phosphoenolpyruvate-dependent
phosphotransferase genes: functional expression in recombinant
Escherichia coli and identification of a putative binding region for
disaccharides."
RL Appl. Environ. Microbiol. 63:355-363(1997).
DR EMBL; U61727; AB51563.1;
DR HSSP; P20166; IGPB.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA.1;
DR Pfam; PF00367; PTS_EIIB.1;
DR Pfam; PF02378; PTS_EIIC.1;
DR ProDom; PD001476; PTS_EIIB.1;
DR ProDom; PD002243; PTS_EIIA.1;
DR PROSITE; PS00371; PTS_EIIA.1;
DR PROSITE; PS00371; PTS_EIIB.CYS.1;
SQ SEQUENCE 621 AA; 65492 MW; DFIDE2A2A7B81339 CRC64;

Query Match 25.7%; Score 468; DB 2; Length 621;
Best Local Similarity 33.0%; Pred. No. 5.3e-21;
Matches 120; Conservative 67; Mismatches 129; Indels 48; Gaps 9;

QY 7 VGGLLFGLVSPVITGLHOSPPPIEL-ELFNQGGSFIFATASMANIAOGAACLAFFFLA 65
DB 290 LAGAAAGALWQVCVIFGLHWGLIPLMNNNAVIGLHMMPLLPVAVMGQVGAALGIFLRT 349
QY 66 KSEKLGAGAGSVAVLITEPAIFGVNLRWRPFIFIGTAAIGGALIALFNKIKAV 125

```
Db 350 RDARQKVLGASAVSAGIFGVTEPAIYGLNPLRRPFIFGCYVWPIGAMVG-FSDSHVST 408
QY 126 GAAGFLGVVS-----IDAPDMWFLVCVVTFEIAFGAAIAYGLYLVRNGSIDPD 176
Db 409 YSFGFNIITLAQWIPPEGIDAT-----VWGGAAGMFASLII----- 445
QY 177 ATAAPVPAGTTKAAE-----APAEFSNDSTIIQAPLTGEAIALSSVSDFAMFASGLKGS 230
Db 446 ACVLTIVAGLPRSSAEQAQAAVWAPASV-ND---ILAPMTGSLVALDQVPDSTFASGLLGQ 501
QY 231 GVAIVPTKGLVSPVSGKIVVAFPSGHAFVARTKAEDGSNVDILMHIGFTDVLNLTGHTFN 290
Db 502 GVAILIISVGVIAFPGFVASIFQTHAIGLLS-----DSGIELLIHVGDITVKLDGAPFT 557
QY 291 PLKKQGVDEKAGELLCEFDIDAKAAGYEVTTPIVVSNYKKTGPVNTYGLGETEAGANLL 350
Db 558 AHVKEGDKIKAGDULLEFFDRAQILDAGYDLATPIIISNSDDFTLDVNSASAVDAQPLL 617
QY 351 NVAK 354
Db 618 SVSR 621

RESULT 11
Q9CFK9
ID Q9CFK9 PRELIMINARY; PRT; 636 AA.
AC Q9CFK9;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE BETA-GLUCOSIDE-SPECIFIC PTS SYSTEM IIABC COMPONENT (EC 2.7.1.69).
GN PTBA.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TL1403;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis."
RL Genome Res. 0:0-0(2001).
DR EMBL; AB006376; AAK05558.1; -.
DR InterPro: IPR003439; ABC transportr.
DR InterPro: IPR001127; PTS_EIIB.
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR ProDom: PD002243; PTS_EIIB; 1.
DR ProSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 636 AA; 67630 MW; 012030D819163325 CRC64;
```

Query Match 24.6%; Score 448; DB 2; Length 636;
Best Local Similarity 33.0%; Pred. No. 9.2e-20;
Matches 110; Conservative 59; Mismatches 136; Indels 28; Gaps 7;

```
QY 3 FGGPVGLLFLGVSPVITGLHQSFPPIELELFNQG-CSFIFATASMANIAQGAACLA 61
Db 288 FNPIIFGLVGRVQWVLMFGLHWAIVPFAIILAKGEPTALLIAASVASFAQTGAVG 347
QY 62 FFLAKSEKLGLAGSGVAVLGITPAIFGVNLRWPFIFIGTAAIGGALIALFNK 121
Db 348 MLKTKDKRLRELAIPAFISGWFGVTEPAIYGITLPKKRPFVWSCIIVSGILSAMALG 407
QY 122 AVNLAGAGFLGVVSDIDAPDMWFLVCVVTFEIAFGAAIAYGL---YLVRNGSIDPD 178
```

```
Db 408 AYTMA---LGIFSFTSNITLTGEVSGAIIKIMIVSAVAVIAGFVVTVL---GFEDDVI 460
QY 179 AAPVPAGTTKAAEAPAEFSNDST---IIQAPLTGEAIALSSVSDFAMFASGLKGSVAIV 235
Db 461 ENPIP-----DKKNKQKTKEIGSPLEGKVIPLSQVKDAFSAFSGVWGKAIE 510
QY 236 PTKGOLVSPVSGKIVVAFPSGHAFVARTKAEDGSNVDILMHIGFTDVLNLTGHTFN 295
Db 511 PTLGEVRAFPDGMVMILFPTKHAAGLIS---NEGTELLIHIGIDTVQLGKGFETVFKQ 566
QY 296 GDEVKAGELLCEFDIDAKAAGYEVTTPIVVSN 328
Db 567 GOSVKKGDLLKFDIERIONAGYSTQVPIVTN 599

RESULT 12
Q9A0X4
ID Q9A0X4 PRELIMINARY; PRT; 620 AA.
AC Q9A0X4;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE BETA-GLUCOSIDE PERMEASE IIABC COMPONENT.
GN SPY0572.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006513; AAK33557.1; -.
KW Complete proteome.
SQ SEQUENCE 620 AA; 66148 MW; 1AF6872CFDD7C7D6 CRC64;
```

Query Match 24.5%; Score 445.5; DB 2; Length 620;
Best Local Similarity 33.0%; Pred. No. 1.3e-19;
Matches 107; Conservative 63; Mismatches 131; Indels 23; Gaps 7;

```
QY 11 LFLGVSPVITGLHQSFPPIELELFNQGSFIFATASMA-NIAQGAACLAIVFLAKSE 69
Db 288 ILGAIMPLVVMVTGMHWAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAFKTKQ 347
QY 70 LKGLAGASGVAVL-GITEPAIFGVNLRWPFIFIGTAAIGGALIALFNKAVAGAA 128
Db 348 TRQVALAAGISALLAGITEPALYGVTLKFKKPLAAMISGGLVGAFIGFVNTIASYTFV 407
QY 129 GLGVVVSIDAP---DMWMLVCVVTFEIAFGAAIAYGLYLVRNGSIDPDATAAPVPA 184
Db 408 SIIGLPQYINPSGGANFTNALIAGTATIVLAFSLTFWFG-----IDEE---SPKQV 455
QY 185 GTTKAAEAPAEFSNDSTIIQAPLTGEAIALSSVSDFAMFASGLKGSVAIVPTKGLVSP 244
Db 456 SVAADMSSQVKSGLSTKQT-LYAPMTGEMFLSEVPDPTFSSKLLGEGFALLPGEVYAP 514
QY 245 VSGKIVVAFPSGHAFVARTKAEDGSNVDILMHIGFTDVLNLTGHTFNPLKKQGVDEK 304
Db 515 FDEGVITFPTTKHAVALK----NTRGVEVLHVIGIDTVLKGQGFQQLVSGVDVVRGQA 570
QY 305 LCEFDIDAKAAGYEVTTPIVVSN 328
Db 571 LLKMDIDFTSKGYSILSPVVVN 594

RESULT 13
```

```
Q9X565
ID Q9X565 PRELIMINARY; PRT; 618 AA.
AC Q9X565;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE BETA-GLUCOSIDE SPECIFIC TRANSPORT PROTEIN.
GN BGLS.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1352;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BFE 900;
RX MEDLINE=99240446; PubMed=10224016;
RA Franz C.M.A.P., Worobo R.W., Quadri L.E.N., Schilling U.,
RA Holzapfel W.H., Vederas J.C., Stiles M.E.;
RT "Atypical genetic locus associated with constitutive production of
RT enterocin B by enterococcus faecium BFE 900.";
RL Appl. Environ. Microbiol. 65:2170-2178(1999).
DR EMBL; AF121254; AAD28228.1; -.
DR HSSP; P20166; IAX3.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR PRODOM; PD002243; PTS_EIIA; 1.
DR PROSITE; PS00371; PTS_EIIA_1; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
SQ SEQUENCE 618 AA; 67051 MW; 9AECDC16BF2DEC9C CRC64;

Query Match 23.8%; Score 433.5; DB 2; Length 618;
Best Local Similarity 35.1%; Pred. No. 6.9e-19;
Matches 128; Conservative 62; Mismatches 130; Indels 45; Gaps 15;

QY 7 VGLLFLGLVYSP---IVITGLHQSPFPIELFLNQGG---SFIFATASMANIAOQAAC 59
DB 278 VGSWFVFMVLYSALQPFITMLGAGNFIMPVVASLIAANGYDPAFI-SSCTISDIAYGGAML 336
QY 60 AVFFLAKSEKLGAGASGVSAVLGITPAIFGVNLRWRPFPIGTAAGGALIALFN 119
DB 337 GYFLTRNLKQKLGFTVLSVAVGTPAIVGVVRRPVPVAVMIGGGLGGLFAGLT 396
QY 120 IKAVALGAAGFLGVVS-IDAPDMV--MFLVCAVTFVFFIAFGAAIAYGLYVLRNGSIDPD 176
DB 397 VKAYSV-AWGLFGLPAYIGEGDFMFWFVLAAVVISFA--GSAVA--AYLL----- 442
QY 177 ATAAPVPAGTTKAEA--EAPAEFSDSTIIQAPL-----TGAEIALSSVSDAMFASGKLGS 230
DB 443 ----GVPAQEEDEQETINIIAEKGDKNLRTSSLGNVVEGEVISLNEVKDQAFSTGALCK 498
QY 231 GVAIVPTKGLVSPSGKIVAFPSGHAFVTKAEKDSNDILMHIGFDTVNLNGTHEN 290
DB 499 GTAIKPIDTHLYSPVDGETVVFPTNHAIGI--QSQDQ--IEILLHIGDITVELEGKYN 554
QY 291 PLKKGDEVKAGELCEFDIDAIAKAAGYEVTTPIVVSN---YKKTGPVNTYGLGIEAGA 347
DB 555 LNIKQDKIQRCOLLGTVDKIQEAGYDTSTIVVVTWNTSDSLVDVIPSNS---EQVIKTD 611
QY 348 NLLNV 352
DB 612 NLLNV 616

RESULT 14
Q99RQ0
ID Q99RQ0 PRELIMINARY; PRT; 480 AA.
AC Q99RQ0;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
```

```
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN SCRA OR SA2167.
OS Staphylococcus aureus subsp. aureus N315.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43469.1; -.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51232 MW; D88607F6E0AF2E84 CRC64;

Query Match 23.0%; Score 418.5; DB 2; Length 480;
Best Local Similarity 49.7%; Pred. No. 4.1e-18;
Matches 89; Conservative 32; Mismatches 47; Indels 11; Gaps 4;

QY 1 YDFGGPVGGLLGLVYSPIVITGLHQSPPIELFLF---NNGSGFIFATASMANIAOQA 56
DB 301 YEFGAIGGLIFGLIYAPVITGMHHSFIAVETTLIADATKGTGGSFIFPIATMSNVAQGG 360
QY 57 ACLAVVFFLAK-SEKLGKLAGASGVSAVLGITPAIFGVNLRWRPFPIGTAAGGAL 115
DB 361 AIAAFTTITKONKLGKVASAAGISALLGITPEPAMFGVNLKLRYPFPIAIVGSGICSAVI 420
QY 116 ALFNKKAVALGAAGFLGVVSDA--PDMVMLVCAVTFVFFIAFGAAIAYGLYVLRNGS 172
DB 421 AFFKVKATALGATAGLPFGFISINPVHAGLWHLVFGVMTISFII---ALTVTLSKRKAN 475

RESULT 15
Q9KLT8
ID Q9KLT8 PRELIMINARY; PRT; 479 AA.
AC Q9KLT8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT.
GN VCA0653.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emcleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004395; AAF96554.1; -.
DR TIGR; VCA0653; -.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:28:49 ; Search time 17.08 seconds

(without alignments)
777.088 Million cell updates/sec

Title: US-09-604-231-4

Perfect score: 1821

Sequence: 1 YDFGPGVGLLGLVLPV.....IEAGANLLNVAKEAVPATP 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	711	39.0	664	1 PTSA_STRMU	P12655 streptococ
2	706.5	38.8	651	1 PTSA_PEDPE	P43470 pediococcc
3	515	28.3	609	1 PTBA_BACSU	P40739 bacillus su
4	492.5	27.0	631	1 PTBA_ERWCH	P26207 erwinia chr
5	450	24.7	625	1 PTBA_ECOLI	RN0722 escherichia
6	446.5	24.5	480	1 PTBS_STAXY	P51184 staphylococ
7	418	23.0	474	1 PTBS_PASMU	Q9C322 pasteurella
8	412	22.6	479	1 PTBS_VIBAL	P22825 vibrio algi
9	386.5	21.2	674	1 PTGA_CORGL	Q45298 corynebacte
10	330	18.1	470	1 PTTB_BACSU	P39794 bacillus su
11	313	17.2	651	1 PTAA_KLEPN	P45604 klebsiella
12	305	16.7	460	1 PTBS_BACSU	P05306 bacillus su
13	299	16.4	699	1 PTGA_BACSU	P20166 bacillus su
14	296.5	16.3	456	1 PTBA_SALTY	P08470 salmonella
15	295	16.2	648	1 PTAA_ECOLI	P09323 escherichia
16	294.5	16.2	324	1 PTGA_BACST	P42015 bacillus st
17	293.5	16.1	456	1 PTBS_KLEPN	P27219 klebsiella
18	273	15.0	459	1 SACK_BACSU	P15400 bacillus su
19	270.5	14.9	631	1 YBFS_BACSU	P39816 bacillus su
20	256	14.1	189	1 PTGA_BORBU	Q44840 borrelia bu
21	255	14.0	168	1 YPOE_BACSU	P50829 bacillus su
22	252.5	13.9	634	1 LACT_STRTR	P23936 streptococ
23	249.5	13.7	154	1 PTGA_MYCCA	P45618 mycoplasma
24	249.5	13.7	473	1 PTTB_ECOLI	P36672 escherichia
25	246	13.5	168	1 PTBA_ECOLI	P08837 escherichia
26	244	13.4	168	1 PTGA_SALTY	P02908 salmonella
27	238.5	13.1	165	1 PTGA_HREIN	P45338 haemophilus
28	231.5	12.7	161	1 PTGA_BUCAI	Q9wx17 buchnera ap
29	221	12.1	641	1 RAFP_PEDPE	P43466 pediococcc
30	219.5	12.1	627	1 LACT_LACDE	P22733 lactobacilli
31	208.5	11.4	940	1 PTGA_MYCPN	P75569 mycoplasma
32	197.5	10.8	908	1 PTGA_MYCGE	P47315 mycoplasma
33	165.5	9.1	485	1 PTDA_ECOLI	P24241 escherichia

34	161.5	8.9	482	1	PTSB_VIBCH	Q9kvd9 vibrio chol
35	157.5	8.6	639	1	LACY_LEULA	Q48624 leuconostoc
36	151.5	8.3	474	1	YFEV_ECOLI	P77272 escherichia
37	115.5	6.3	538	1	TCMA_STRGA	P39886 streptomyce
38	113.5	6.2	1179	1	CIAD_BACTA	Q03744 bacillus th
39	110	6.0	1951	1	CIN3_HUMAN	Q9ny46 homo sapien
40	109	6.0	414	1	FABF_VIBHA	P55338 vibrio harv
41	107.5	5.9	2249	1	OMPA_RICRI	P15921 rickettsia
42	104	5.7	451	1	PTCC_BACST	Q45400 bacillus st
43	104	5.7	577	1	PTLB_LACCA	P24400 lactobacilli
44	101.5	5.6	413	1	FABF_VIBCH	Q9kqn9 vibrio chol
45	101.5	5.6	1654	1	OMPB_RICRI	Q53047 r outer mem

ALIGNMENTS

RESULT	1
PTSA_STRMU	STANDARD; PRT; 664 AA.
AC	P12655;
DT	01-OCT-1989 (Rel. 12, Created)
DT	01-OCT-1989 (Rel. 12, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (ETIABC-SCR) (SUCROSE-
DE	PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE	COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN	SCRA.
OS	Streptococcus mutans.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1309;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=GS-5;
RC	MEDLINE=89123027; PubMed=2536656;
RA	Sato Y., Poy F., Jacobson G.R., Kuramitsu H.K.;
RT	"Characterization and sequence analysis of the scra gene encoding
RT	enzyme IIScr of the Streptococcus mutans
RT	phosphoenolpyruvate-dependent sucrose phosphotransferase system.";
RL	J. Bacteriol. 171:263-271(1989).
RL	[2]
RP	SEQUENCE OF 639-664 FROM N.A.
RC	STRAIN=GS-5;
RC	MEDLINE=93329360; PubMed=8336109;
RA	Sato Y., Yamamoto Y., Kizaki H., Kuramitsu H.K.;
RT	"Isolation, characterization and sequence analysis of the scrK gene
RT	encoding fructokinase of Streptococcus mutans.";
RL	J. Gen. Microbiol. 139:921-927(1993).
CC	-!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC	SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC	-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC	AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC	PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC	PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC	THE SUGAR.
CC	-!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR -
CC	PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC	-!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC	-!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC	-!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; M22711; AAA26971.1; -
DR	EMBL; D13175; BAA02466.1; -


```
DR PIR; B32243; B32243.
DR HSP; P08837; 2F3G.
DR InterPro: IPR001127; PTS_EIIA.
DR InterPro: IPR00196; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF003358; PTS_EIIA.1; 1.
DR Pfam: PF003367; PTS_EIIB.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR ProDom: PD002243; PTS_EIIA; 1.
DR ProSite: PS00371; PTS_EIIA.1; 1.
DR ProSite: PS01035; PTS_EIIB; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40
FT DOMAIN 2 64
FT MOD_RES 25
FT MOD_RES 331
FT MOD_RES 585
SQ SEQUENCE 664 AA; 69988 MW; 809B63E32281A9A1 CRC64;

Query Match 39.0%; Score 711; DB 1; Length 664;
Best Local Similarity 42.1%; Pred. No. 3.8e-43;
Matches 155; Conservative 73; Mismatches 116; Indels 24; Gaps 8;

QY 1 YDFGGPVGGLLFGVLVSPVITGLHQSFPPIELLEL---FNQG---GSFIATASMANIAQ 54
Db 307 YDTGFLMGVFGALYSPVVTGLHQSFPPIELLEL---FNQG---GSFIATASMANIAQ 366

QY 55 GAACLAFFFLAKSEKLGAGAGSVAVLGTETPAIFGVNLRWLPFFIGTAAIGCAL 114
Db 367 GAATFAIYELTKDKMKGLSSSGVSALLGTETPAIFGVNLRWLPFFIGTAAIGCAL 426

QY 115 IALENIKAVLGAAGFLGVVSDADPMVFLVCAVVTFFAFAAGIAYGLVLRNNGSID 174
Db 427 AGLLQVAVSLGSAFGFLGSLKASSIPFYVCELISFAIAFAVYGYG----KTKAVD 481

QY 175 PDATAAPVPAGTAKAEAEAPFSDST-----IIQAPLTGAIALLSSVSDAMFASGK 227
Db 482 VFRAEAEEAEIEVQ-EIPEEAASAANKAQVDTDEVLAAFLAGEAVELTSVNDPVFSEA 540

QY 228 LGSVAIVPTKQGLVSPVSKIVVAFPSGHAFVTRKAEDGNSVDILMHGFDTVNLNGT 287
Db 541 MGKGIATKPSGNTVYAPVDGTQVAFDTGHAYGI--KSDNGA--EILIHIGIDTVSMEGK 596

QY 288 HFNPCLKQGEVKAELLCEFDIDAIAKAGYEVTTPIVVSNNYKKGTPVNTY-GLGEIEAG 346
Db 597 GFQKQVADQKIRKGDVLGTGFDSDKIAEAGLDNTTMEFIVNTADYASVETLASSGTAVAG 656

QY 347 ANLINVAK 354
Db 657 DLSLEVKK 664

RESULT 2
PTSA_PEDPE
ID PTSA_PEDPE STANDARD; PRT; 651 AA.
AC P43470;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC ITABC COMPONENT (EIIABC-SCR) (SUCROSE-
DE PERMEASE ITABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Pedicoccus pentosaceus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Pedicoccus.
OX NCBI_TaxID=1255;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=PEP1.0;
RA Leenbouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXI-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING-SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z32771; CAA83668.1; -.
CC HSP; P20166; IAX3.
CC InterPro: IPR001127; PTS_EIIA.
CC InterPro: IPR00196; PTS_EIIB.
CC InterPro: IPR003352; PTS_EIIC.
CC Pfam: PF003358; PTS_EIIA.1; 1.
CC Pfam: PF003367; PTS_EIIB; 1.
CC Pfam: PF02378; PTS_EIIC; 1.
CC ProDom: PD001476; PTS_EIIB; 1.
CC ProDom: PD002243; PTS_EIIA; 1.
CC ProSite: PS00371; PTS_EIIA.1; 1.
CC ProSite: PS01035; PTS_EIIB; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
KW DOMAIN 1 40
KW DOMAIN 2 64
KW MOD_RES 25
KW MOD_RES 324
KW MOD_RES 562
SQ SEQUENCE 651 AA; 68454 MW; C87BA09D550A77F8 CRC64;

Query Match 38.8%; Score 706.5; DB 1; Length 651;
Best Local Similarity 42.0%; Pred. No. 7.6e-43;
Matches 154; Conservative 64; Mismatches 126; Indels 23; Gaps 5;

QY 1 YDFGGPVGGLLFGVLVSPVITGLHQSFPPIELF-----NQGSFIFATASMANIAQGA 56
Db 300 YNSTGWTGCMGIFGLLYSAIVITGLHQTTPAETQTLLANVAKTGSFIFPVASMANIQGA 359

QY 57 ACLAVFFLAKSEKLGAGAGSVAVLGTETPAIFGVNLRWLPFFIGTAAIGGALIA 116
Db 360 ATLAIFFATKSOKKALTSSAGVSALLGTETPAIFGVNLRWLPFFIFATASGIAFLG 419

QY 117 LFNIKAVLGAAGFLGVVSDADPMVFLVCAVVTFFIAGAAIAYGLYLVRNNGSIDPD 176
Db 420 LFHVLVAMGPASVIGFISTASKSIAPAFMLSAVISFVVAIFPIFY----AKRTLGDDR 475

QY 177 ATAAPVPAGTAKAEAEAPFSDSTIIQAPLTGAIALLSSVSDAMFASGKLGVAIVP 236
Db 476 QVKSAPTSTV-----INVNDEIISAPVTGASESLKQVNDQVESAIEIMKGAAIVP 526

QY 237 TKGLVSPVSGKTLVAFPPSGHAFVTRKAEDGNSVDILMHGFDTVNLNCTHENPLKKG 296
Db 527 SSDQVAVAPDGVITVTDYDSSHAYGIKTTA-----GAELIHLGLDVTNVLNGEHFTTNNVQK 582
```


QY 357 AVPA 360
 DB 604 ALLA 607

RESULT 4
 PTBA_ERWCH STANDARD; PRT; 631 AA.

AC P26207;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIAABC COMPONENT (EIIABC-BGL)
 DE (BETA-GLUCOSIDES-PERMEASE IIAABC COMPONENT) (PHOSPHOTRANSFERASE
 DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).
 GN ARBF.
 OS Erwinia chrysanthemi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92121114; PubMed=1732212;
 RA el Hassouni M., Henrissat B., Chippaux M., Barras F.;
 RT "Nucleotide sequences of the arb genes, which control beta-glucoside
 RT utilization in Erwinia chrysanthemi: comparison with the Escherichia
 RT coli bgl operon and evidence for a new beta-glycohydrolase family
 RT including enzymes from eubacteria, archaeobacteria, and humans";
 RL J. Bacteriol. 174:765-777(1992).
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.
 CC -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY
 CC SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81772; AAA24814.1;
 DR PIR; B42603; B42603.
 DR HSP; P20166; IAX3.
 DR InterPro; IPR001127; PTS_EIIA.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00358; PTS_EIIA_1; 1.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR ProDom; PD002243; PTS_EIIA; 1.
 DR ProSITE; PS00371; PTS_EIIA_1; 1.
 DR ProSITE; PS01035; PTS_EIIB_Cys; 1.
 KW Phosphotransferase system; Sugar transport; Transferase;
 KW Phosphorylation; Transmembrane; Inner membrane.
 FT DOMAIN 1 41 EIIB DOMAIN.
 FT DOMAIN ? ? EIIC DOMAIN.
 FT DOMAIN 501 631 EIIA DOMAIN.
 FT MOD_RES 26 26
 FT MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
 FT PHOSPHORYLATION (BY SIMILARITY).

FT MOD_RES 553 553 PHOSPHORYLATION (BY SIMILARITY).
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 146 166 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT TRANSMEM 206 226 POTENTIAL.
 FT TRANSMEM 248 268 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 295 315 POTENTIAL.
 FT TRANSMEM 328 348 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 407 427 POTENTIAL.
 FT TRANSMEM 434 454 POTENTIAL.
 SQ SEQUENCE 631 AA; 66984 MW; 7D0BD27A36BFFDF CRC64;

Query Match 27.0%; Score 492.5; DB 1; Length 631;
 Best Local Similarity 34.9%; Pred. No. 1e-27;
 Matches 123; Conservative 62; Mismatches 154; Indels 13; Gaps 7;

QY 1 YDFGGVGGLLFGLVSPVITGLHOSFPPIELEFNO--GGSFIFATASMANIAQGAACL 59
 DB 285 YGLNSLLAGAVMGALQVQCVIFGLHWGFVPLMLNNSFVIGHDTLLPLLVPAVLQAGATL 344
 QY 60 AVFLAKSEKLKGLAGASGSAVLGITEPAIFGVNLRWPRFFIGITGAIGALLALFN 119
 DB 345 GVLLRTQDLKRKGIAAGSAFALGITEPAVIGVTPLRRPFIFGCGALGAAGVGYAH 404
 QY 120 IKAVALGAAGFLGVWSIDAPDMVFLV--CAVYTFITAFG--AAIAYGLYLVRNRGSDPDA 177
 DB 405 TTMYSFGFFSIESTQVPIPTGVDSSVMAAVIGTLLAFALTSNFGVPKD---ETQP 461
 QY 178 TRAPVPAGTTKAEAPAEFNSNDSTIIQAPLGEATLSSVSDAMFASGKLGSGVAIVPT 237
 DB 462 AAADSPA--VLAETQANAGAVRDETLLF--SPLAGEVLLLEQVADRTTFASGVMKGIAIRPT 518
 QY 238 KGLVSPVSGKIVVAFPSGHAFARVTKAEDGSNVDILMHIGEDTVNLNGTHRNPLKKG 297
 DB 519 QRLYAPVGTVASLKTKTHAIGLASR---GGAEVLIHVGIDTVRLDGRYTFPHVRVGD 574
 QY 298 EVKAGELLCFFDIDAIAKAGYEVTTPIVVSNYKKTKGTPVNTYGLGEIEAGANL 349
 DB 575 VVRQGLLLLEFDGPAIEAAGYDLTTPITVNTSEYDYGVEPVASGKVDANAPL 626

RESULT 5
 PTBA_ECOLI STANDARD; PRT; 625 AA.

AC P08722;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PTS SYSTEM, BETA-GLUCOSIDES-SPECIFIC IIAABC COMPONENT (EIIABC-BGL)
 DE (BETA-GLUCOSIDES-PERMEASE IIAABC COMPONENT) (PHOSPHOTRANSFERASE
 DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-BGL).
 GN BGLF OR BGLC OR BGLS OR B3722.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=88009877; PubMed=3309161;
 RA Branley H.F., Kornberg H.L.;
 RT "Nucleotide sequence of bglC, the gene specifying enzymelbgl of the
 RT PEP:sugar phosphotransferase system in Escherichia coli K12, and
 RT overexpression of the gene product";
 RL J. Gen. Microbiol. 133:563-573(1987).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12;
 RX MEDLINE=87222180; PubMed=3034860;
 RA Schnetz K., Toloczkyi C., Rak B.;
 RT "Beta-glucoside (bgl) operon of *Escherichia coli* K-12: nucleotide
 sequence, genetic organization, and possible evolutionary
 relationships to regulatory components of two *Bacillus subtilis*
 genes";
 RT J. Bacteriol. 169:2579-2590(1987).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RC MEDLINE=93315143; PubMed=7686882;
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
 RT "DNA sequence and analysis of 136 kilobases of the *Escherichia coli*
 genome: organizational symmetry around the origin of replication";
 RL Genomics 16:551-563(1993).
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 -TRANSPORT SYSTEM. THE IIDC DOMAINS CONTAIN THE SUGAR BINDING SITE
 AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 THE SUGAR.
 CC -!- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON BGLG.
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M15746; AAA83837.1; -;
 DR EMBL; M16487; AAA23510.1; -;
 DR EMBL; L10328; AAA62073.1; -;
 DR EMBL; AE000449; AAC76745.1; -;
 DR PIR; C25977; C25977.
 DR PIR; A47616; A47616.
 DR HSP; P20166; IAX3.
 DR Ecogene; EGI0115; bglF.
 DR InterPro; IPR001127; PTS_EIIA.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00358; PTS_EIIA_1; 1.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR ProDom; PD002243; PTS_EIIA; 1.
 DR ProSITE; PS00371; PTS_EIIA_1; 1.
 DR ProSITE; PS01035; PTS_EIIB_CYS; 1.
 KH Phosphotransferase system; Sugar transport; Transferase;
 KW Phosphorylation; Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 1 39
 FT DOMAIN ? ?
 FT DOMAIN 495 625
 FT MOD_RES 24 24
 FT MOD_RES 306 306
 FT MOD_RES 547 547
 FT TRANSMEM 59 79
 FT TRANSMEM 100 120
 FT TRANSMEM 141 161
 FT TRANSMEM 167 187
 FT TRANSMEM 203 223
 FT TRANSMEM 245 265
 FT TRANSMEM 285 305
 FT TRANSMEM 325 345

FT TRANSMEM 355 375
 FT TRANSMEM 381 401
 FT TRANSMEM 404 424
 FT TRANSMEM 432 452
 FT TRANSMEM 490 510
 FT TRANSMEM 526 546
 SQ SEQUENCE 625 AA; 66482 MW; 9E12A2207125C4E6 CRC64;
 Query Match 24.7%; Score 450; DB 1; Length 625;
 Best Local Similarity 33.3%; Pred. No. 1.le-24;
 Matches 116; Conservative 64; Mismatches 154; Indels 14; Gaps 6;
 Qy 9 GLIFGLVSPVITGLHQSFPPIELELEN-QGSGFIFATASMANIAOGAACIAVFEFLAKS 67
 Db 290 GAVMGGFQWIFVMFGLHGLVPLCINNFVTDYDTMIPMLPAIMAVQGAALGVFLCERD 349
 Qy 68 EKLKLAGASGVSAVLGITEPAIFGNVLRMRPFPIGIGTAAIGGALIALFNKAVALGA 127
 Db 350 AOKKVAGSAALTSFGITEPAVGVNLPKPYFVACISGALGATIIGYAQTQVYFSL 409
 Qy 128 AGFLGVSDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRNGSIDPD-ATAAPVPAGT 186
 Db 410 PSIFTFMOTIPSTGIDFTWASV----IGGVIAIGCAFV---GTVMLHFIATKRPQAG 461
 Qy 187 TKAEAEAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKQOLVSPVS 246
 Db 462 APOKTEPVITPEOGGICSPMTGEIVPLIHVADTFEASGLLGKGLAILPSVGEVRSVA 521
 Qy 247 KIVVAFPSGHFAVTRKADGNSVDILMHIGFDVNLNGTHFNPLKKGQDEVKAGELIC 306
 Db 522 GRIASLFATLHAIGI--ESDDG--VEILHVGIDIVKLDGKFFSAHVNVGDKVNTGDRLI 577
 Qy 307 EFDIDAIKAAGVEVTPVIVSNYKKTGPNYTVGLGEIEAGANLNVAK 354
 Db 578 SFDIPAIREAGFDLTPTVLISNSDDFTDVLPHGTAQISAGEPLLSIIR 625
 RESULT 6
 PTSE_STAXY STANDARD; PRT; 480 AA.
 ID PTSE_STAXY
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIBC-SCR) (SUCROSE-
 PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
 DE (EC 2.7.1.69) (EII-SCR).
 GN SCRA.
 OS Staphylococcus xyloso.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1288;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 20267 / C2A;
 RX MEDLINE=94049686; PubMed=8232209;
 RA Wagner E., Goetz F., Brueckner R.;
 RT "Cloning and characterization of the *scrA* gene encoding the sucrose-
 specific Enzyme II of the phosphotransferase system from
 Staphylococcus xyloso.";
 RL Mol. Gen. Genet. 241:33-41(1993).
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 -TRANSPORT SYSTEM. THE IIDC DOMAINS CONTAIN THE SUGAR BINDING SITE
 AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 THE SUGAR.
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.

```

CC CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X69800; CAA9461.1; -.
CC CC InterPro; IPR001996; PTS_EIIB.
CC CC InterPro; IPR003352; PTS_EIIC.
CC CC Pfam; PF00367; PTS_EIIB; 1.
CC CC Pfam; PF02378; PTS_EIIC; 1.
CC CC ProDom; PD001476; PTS_EIIB; 1.
CC CC PROSITE; PS01035; PTS_EIIB_CYS; 1.
CC CC Phosphotransferase system; Sugar transport; Transferase;
CC CC Transmembrane; Inner membrane; Phosphorylation.
CC CC DOMAIN 1 ? EIIB DOMAIN.
CC CC FT DOMAIN ? 480
CC CC FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
CC CC FT MOD_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).
CC CC SEQUENCE 480 AA; 51326 MW; AB4E1D9785D84E47 CRC64;

Query Match 24.5%; Score 446.5; DB 1; Length 480;
Best Local Similarity 52.5%; Pred. No. 1.4e-24;
Matches 93; Conservative 30; Mismatches 47; Indels 7; Gaps 3;

QY 1 YDFGGPVGGLFLGVYSPVITGLHQSFPPIELELF-----NOGGSFIFATASMANIAOQA 56
Db 301 YEFGAIGGLIFGLLYAPIVTGMHHSFIAETOLIADSSSTGSGSFIPIATMSNIAOQA 360
QY 57 ACLAVFFFLAK-SEKLGKLAGAGSVAVLIGTEPAIFGVNLRWPPFIFIGTAAIGGALI 115
Db 361 AALAAFFIIKENKLGKVASAAGVSALLGITPEPAMEGVNLRYPFIFIGTAAIGGALI 420
QY 116 ALFNKAVAGLGAAGFLGVVSIDAPD--MYMFLVCVVTFEIAFGRAIYGLVLRN 170
Db 421 APEKVAIALGATAGIPGFTISGONGWUHYGIAMIIAFIVAGVTYALSYRKKYRN 477

RESULT 7
PTS_PASMU STANDARD; PRT; 474 AA.
AC Q9CJ22.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIB-SCR) (SUCROSE-
DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EI-SCR).
GN SCRA OR PTSB OR PM1846.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pn70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =

```

```

CC CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC CC (BY SIMILARITY).
CC CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AE006222; AAK03930.1; -.
CC CC InterPro; IPR001996; PTS_EIIB.
CC CC InterPro; IPR003352; PTS_EIIC.
CC CC Pfam; PF00367; PTS_EIIB; 1.
CC CC Pfam; PF02378; PTS_EIIB; 1.
CC CC PROSITE; PS01035; PTS_EIIB_CYS; FALSE_NEG.
CC CC Phosphotransferase system; Transferase; Sugar transport; Plasmid;
CC CC Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
CC CC DOMAIN 1 ? EIIB DOMAIN.
CC CC FT DOMAIN 109 474
CC CC FT TRANSMEM 109 129 POTENTIAL.
CC CC FT TRANSMEM 158 178 POTENTIAL.
CC CC FT TRANSMEM 182 202 POTENTIAL.
CC CC FT TRANSMEM 229 249 POTENTIAL.
CC CC FT TRANSMEM 284 284 POTENTIAL.
CC CC FT TRANSMEM 303 323 POTENTIAL.
CC CC FT TRANSMEM 345 365 POTENTIAL.
CC CC FT TRANSMEM 376 396 POTENTIAL.
CC CC FT TRANSMEM 403 423 POTENTIAL.
CC CC FT TRANSMEM 444 464 POTENTIAL.
CC CC FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
CC CC FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
CC CC SEQUENCE 474 AA; 49908 MW; D26FA1A059603C01 CRC64;

Query Match 23.0%; Score 418; DB 1; Length 474;
Best Local Similarity 50.9%; Pred. No. 1.4e-22;
Matches 85; Conservative 31; Mismatches 47; Indels 4; Gaps 1;

QY 1 YDFGGPVGGLFLGVYSPVITGLHQSFPPIELELFNQ-----GGSFIFATASMANIAOQA 56
Db 300 YDTLGVGGAIFGLTYAPIVTGMHQTFAVETQLAEVARTGCTFIPIAAMSNIAOQA 359
QY 57 ACLAVFFFLAKSEKLGKLAGAGSVAVLIGTEPAIFGVNLRWPPFIFIGTAAIGGALIA 116
Db 360 ACLGAAYVMKDAKVRGIAVPSGISALLGITPEPAMEGVNLRYPFISAMIGAGISSAVIA 419
QY 117 LFNKAVAGLGAAGFLGVVSIDAPDVMYFVLCVVTFEIAFGRAIYNG 163
Db 420 LFNKAVIALGAAGLPGIPSIKIPDLSLAMYCVGMLISASIAFTLTVLIG 466

RESULT 8
PTS_PASMU STANDARD; PRT; 479 AA.
ID PTSB_VIBAL
AC P22825;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIB-SCR) (SUCROSE-
DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EI-SCR).
GN SCRA.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071601; PubMed=2174811;

```

Blatch G.L., Scholle R.R., Woods D.R.;
 "Nucleotide sequence and analysis of the *Vibrio alginolyticus* sucrose
 uptake-encoding region.";
 RL Gene 95:17-23(1990).
 RN [2]
 RP SEQUENCE OF 1-12 FROM N.A.
 RX MEDLINE=91285433; PubMed=2060795;
 RA Blatch G.L., Woods D.R.;
 "Nucleotide sequence and analysis of the *Vibrio alginolyticus* scr
 repressor-encoding gene (scrR).";
 RL Gene 101:45-50(1991).
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
 AND THE TRANSMEMBRANE CHANNEL, THE IIA DOMAIN CONTAINS THE PRIMARY
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 THE SUGAR.
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M76768; AAA27555.1; -;
 DR EMBL; M35009; AAA27557.2; -;
 DR PIR; JQ0781; JQ0781.
 DR HSSP; P05053; IIBA.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR PROSITE; PS01035; PTS_EIIB-CYS; 1.
 KW Phosphotransferase system; Sugar transport; Transferase;
 KW Transmembrane; Inner membrane; Phosphorylation.
 FT DOMAIN 1 ? EIIB DOMAIN.
 FT DOMAIN ? 479 EIIC DOMAIN.
 FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 479 AA; 49890 MW; FBF906B5170E3EB7 CRC64;

 Query Match 22.6%; Score 412; DB 1; Length 479;
 Best Local Similarity 49.7%; Pred. No. 3.8e-22;
 Matches 84; Conservative 32; Mismatches 47; Indels 6; Gaps 3;
 Qy 1 YDFGGPVGGLRFLVYSPVITGLHOSPPPIELF-----NOGGSFIFATASMANIAQA 56
 Db 300 YNTAGVGVGFLIYAPFVITGMHSHFATIAETQLADIATTGGTFIPPIAAMSNVSOQA 359
 Qy 57 ACLAVFFFLAKSEKLGAGASGVAVLITEPAIFGVNLRWLPFFIGITAAIGGALIA 116
 Db 360 AALAVGVMSKKKMGKGAIPSGVTGLLGITEPAMGVNLRKLYPFIAVCAALSSAFIT 419
 Qy 117 LFNKAVAGAAAGFVGVSIDAPDMWMLVCA-VVTFITAFGAAYGL 164
 Db 420 MENVKAQALGAAGLGLIISI-TPDKIGYVYIAGWIAFLTAFLVTLVGI 467
 RESULT 9
 FTGA_CORGL
 ID PTGA_CORGL STANDARD; PRT; 674 AA.
 AC Q45298;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PTS SYSTEM, GLUCOSE-SPECIFIC IIB COMPONENT (EIIBC-GLC) (GLUCOSE-
 DE PERMEASE IIB COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
 DE COMPONENT) (EC 2.7.1.69) (EII-GLC/EIIB-GLC).
 GN PTSG.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OG Plasmid pBSBG2.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.
 OX NCBI_TaxID=17118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 13869;
 RA Yoon K.-H.;
 RT "Cloning and nucleotide sequence of enzyme II of Brevibacterium
 lactofermentum phosphotransferase system.";
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
 AND THE TRANSMEMBRANE CHANNEL, THE IIA DOMAIN CONTAINS THE PRIMARY
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 THE SUGAR.
 CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC -!- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L18875; AAA22992.1; -;
 DR HSSP; P08837; 2F3G.
 DR InterPro; IPR001127; PTS_EIIB.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00358; PTS_EIIB; 1.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR ProDom; PD002243; PTS_EIIB; 1.
 DR PROSITE; PS00371; PTS_EIIB; 1.
 DR PROSITE; PS01035; PTS_EIIB-CYS; 1.
 KW Phosphotransferase system; Sugar transport; Transferase;
 KW Phosphorylation; Transmembrane; Plasmid.
 FT DOMAIN 1 43 EIIB DOMAIN.
 FT DOMAIN ? ? EIIC DOMAIN.
 FT DOMAIN 542 674 EIIB DOMAIN.
 FT TRANSMEM 126 146 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 193 213 POTENTIAL.
 FT TRANSMEM 225 245 POTENTIAL.
 FT TRANSMEM 260 280 POTENTIAL.
 FT TRANSMEM 303 323 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 377 397 POTENTIAL.
 FT TRANSMEM 409 429 POTENTIAL.
 FT TRANSMEM 442 462 POTENTIAL.
 FT TRANSMEM 493 513 POTENTIAL.
 FT MOD_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 594 594 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 674 AA; 71626 MW; 67A75AAF76E42FA2 CRC64;

Query Match 21.2%; Score 386.5; DB 1; Length 674;
Best Local Similarity 31.0%; Pred. No. 3.6e-20;
Matches 116; Conservative 61; Mismatches 156; Indels 41; Gaps 9;

QY 10 LFLGLVSPVITGLHQSFPPIELEFNQ--GGSFIFATASMANIAOAGAACLAFFVLAKSE 68
DB 307 IVIPLLYPLVPLGLHPLNAINQNLTLGYDFIQGPMGANFACFGLVGTGFLIALKE 366
QY 69 KLKGLAGAS--GVSAYL--GITEPAIFGVNLRWRPFFTGIGTAATGALIALFNKAVA 124
DB 367 KNAMRQVSLGGLAGLGGISEPSLYGLVLLRFFKTYFRLLPCLLGGVGVIMGFIDIKAYA 426
QY 125 LGAAGLGVVSDAPDMWFLVCAVVTFTAFCAATAYGLYLVRNGSID----- 174
DB 427 F---VFTSLTLLTPAMPDPLWGLYTVGIAAAP--FTSMLLVLFFDYRSDAERDEAKAQAAMAE 481
QY 175 -----PDATAAPV-----PAGTTKAAEAFAEFSNDSTIIQAPLITGBAIALSSV 218
DB 482 QTNNTPAAPAAPVAPAAAGAAAGGAAGAT-AVATKPRLAAGQLVEITSPLEGHAVPLSEV 540
QY 219 SDAMFASGLKSGVAIVPTKQLVSPVSGKIVVAFPSGHAFVTRTAEDGSNVVDILMHIG 278
DB 541 PDPIFAAGLPGIAEPTGTNTVAPADATVILVQKSGHVALRLRLE-----SGVELLIHIG 596
QY 279 FDTVNLNGTHFNPLKKGDEVKAGELLCEDIDAIRKAAAGVEVTTPIWVSNYKKTGPVNTY 338
DB 597 LDTVQLGGEGFKVHVVERKQVQKAGDPLITFDPEFIRSKNPLITIPVVSNAKFGELVGI 656
QY 339 GLGEIEAGANLLNV 352
DB 657 ERAQADATTVIKV 670

RESULT 10
ID PTTB_BACSU STANDARD; PRT; 470 AA.
AC P39794; 034771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE)
DE (TREHALOSE-PECEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC
DE COMPONENT) (EC 2.7.1.69) (EII-TRE).
GN TREP OR TREB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=97074649; PubMed=8917076;
RA Schoeck F., Dahl M.K.;
RT "Analysis of DNA flanking the treA gene of Bacillus subtilis reveals
RT genes encoding a putative specific enzyme II tre and a potential
RT regulator of the trehalose operon.";
RL Gene 175:59-63(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AC327;
RX Yamamoto H., Uchiyama S., Sekiguchi J.;
RT "Cloning and sequencing of a 40.6 kb segment in the 73 degrees-76
RT degrees region of the Bacillus subtilis chromosome containing genes
RT for trehalose metabolism and acetoin utilization.";
RL Microbiology 142:3057-3065(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AC327;
RX MEDLINE=97417488; PubMed=9272861;
RA Yamamoto H., Uchiyama S., Nugroho F.A., Sekiguchi J.;
RT "Cloning and sequencing of a 35.7 kb in the 70 degree-73 degree region

of the Bacillus subtilis genome reveal genes for a new two-component
system, three spore germination proteins, an iron uptake system and a
general stress response protein.";
Gene 194:191-199(1997).
[4]
RP SEQUENCE OF 324-470 FROM N.A.
RC STRAIN=168;
RX MEDLINE=95379486; PubMed=7651129;
RA Helfert C., Gotsche S., Dahl M.K.;
RT "Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by
RT a phospho-alpha-(1-1)-glucosidase encoded by the treA gene.";
RL Mol. Microbiol. 16:111-120(1995).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z54245; CAA91014.1; -;
DR EMBL: D83967; BAA23409.1; -;
DR EMBL: D86417; BAA22289.1; -;
DR EMBL: X80203; CAA56494.1; -;
DR EMBL: Z99108; CAB12609.1; -;
DR HSSP: P05053; 1IBA.
DR Subtilist; BG11009; trep.
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR PROSITE: P001035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Inner membrane; Phosphorylation; Complete proteome.
FT DOMAIN 1 ?
FT DOMAIN ?
FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 301 321 POTENTIAL.
FT TRANSMEM 326 346 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 403 423 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
FT CONFLICT 140 140 F -> S (IN REF. 1).
FT CONFLICT 363 363 M -> L (IN REF. 1 AND 4).
FT CONFLICT 465 465 A -> G (IN REF. 1 AND 4).
SQ SEQUENCE 470 AA; 49999 MW; 7A741850A2697D53 CRC64;

Query Match 18.1%; Score 330; DB 1; Length 470;
Best Local Similarity 41.6%; Pred. No. 2.4e-16;
Matches 67; Conservative 34; Mismatches 58; Indels 2; Gaps 1;

QY 7 VGLLGLVSPVITGLHQSFPPIELEFNQ--GGSFIFATASMANIAOAGAACLAFFVL 64

Db 307 LGGLYGGFYSAVITGMHTTFLAVDLQIGSKGTFLWPMALSNIAQSGSAALAMFWI 366
 QY 65 AKSEKLGAGSAGSVAVLIGTEPAIFGVLNRLRPPFFIGITGTAAGGALIALFNKAVA 124
 Db 367 VKDERQKGLSTGTSAYLIGTEPAIFGVLNRLRPPFFIIVAMVSSGLAGNYISSQGVLA 426
 QY 125 LCAAGFLGVVSDPDMVFLCAVVTFFIAFGAALAYGLY 165
 Db 427 VGVGVPGIFGFSMSQYGAFAIGMAIVLIVPAGTYAYARF 467

RESULT 11
 PTAA_KLEPN
 ID PTAA_KLEPN STANDARD; PRT; 651 AA.
 AC P45604;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIA BC COMPONENT (EIIABC-NAG)
 DE (N-ACETYLGLUCOSAMINE-PERMEASE IIA BC COMPONENT) (PHOSPHOTRANSFERASE;
 DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-NAG).
 GN NAG.

OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=1033-5p14 / KAY2026;
 RX MEDLINE=92079906; PubMed=1745234;
 RA Vogler A.P., Lengeler J.W.;
 RT "Comparison of the sequences of the nagE operons from Klebsiella
 RT pneumoniae and Escherichia coli K12: enhanced variability of the
 RT enzyme IIN-acetylglucosamine in regions connecting functional
 RT domains.";
 RL Mol. Gen. Genet. 230:270-276(1991).

CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.

CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOTIDINE + SUGAR =

CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.

CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.

CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X63289; CAA44923.1; -;
 CC HSSP; P08837; 2F3G.
 DR InterPro; IPR001127; PTS_EIIA.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF00358; PTS_EIIA.1; 1.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR ProDom; PD002243; PTS_EIIA; 1.
 DR PROSITE; PS00371; PTS_EIIA.1; 1.
 DR PROSITE; PS01035; PTS_EIIB.CYS; 1.
 KW Phosphotransferase system; Sugar transport; Transferase;
 KW Phosphorylation; Transmembrane; Inner membrane.

FT DOMAIN 1 ?
 FT DOMAIN 393 427 EIIA DOMAIN.
 FT DOMAIN 519 651 EIIA DOMAIN.
 FT MOD_RES 190 190 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 412 412 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 571 571 PHOSPHORYLATION (BY SIMILARITY).
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 165 185 POTENTIAL.
 FT TRANSMEM 192 212 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 FT TRANSMEM 285 305 POTENTIAL.
 FT TRANSMEM 308 328 POTENTIAL.
 FT TRANSMEM 339 359 POTENTIAL.
 SQ SEQUENCE 651 AA; 68179 MW; 3D8D8ADF4BD48ED CRC64;

Query Match 17.2%; Score 313; DB 1; Length 651;
 Best Local Similarity 23.1%; Pred. No. 5.5e-15;
 Matches 113; Conservative 67; Mismatches 144; Indels 166; Gaps 15;

QY 5 GPGVGLLGLVSPVITGLHOSFPPI-----EELFNOG----GSF 42
 Db 170 GALGSGIFGNRLIPTGLHQVNTIAWFOIGEFTNAACTVPHGDINRFYAGDTAGMF 229
 QY 43 IFATASMANIAQGAACLAFLFAKSEK---LKLAGASGVSAVL-GITEPAIF----- 91
 Db 230 MSGFFPINMFLGPGAALAMYLAAKARRPMVGMGLLSVAITFLTGTETPLEFLFLAP 289
 QY 92 -----GVNLRURWPFFIGITAAIGGAL--IALFNKAVAGAGFLGVVSDAP 139
 Db 290 LLYLLHAVLTGISLTATALGIHAGFSFAGADYVLMYSLPAA-----SK 335
 QY 140 DMVMEVLCVAVTFFI---AFGAIAV-----GL----- 164
 Db 336 NVMLLVGMGVFFVFFVLLFSVIRFMNLTGREDKAADVVTTEANSNTEGLTQIATS 395
 QY 165 YLVRRNG-----SID----- 174
 Db 396 YIAVAGGTDLNKAIDACITRLTLTGDSAKVNDAAACKRLGASGVVKNKOTIQVIVGAKA 455
 QY 175 -----PDATAAPVPAG---TTKAAEAAPAEFSNDSTI--IOAPLTGEAIALS 216
 Db 456 ESTGDEMKVWTRGPVAAAAAAGNVAATAAPAAKQAVANAKTVESLVSPTGDVVALE 515
 QY 217 SYSDAMFASCKLGSVAIVPTKQLVSPVSGKIVVAFPSGHAFVTRKAEDGSNVDLIMH 276
 Db 516 QVPDEAFASKAVGDDGIAVKPTDNIVVAPAGTVVKIFNTNHAFCLET-----NNGAEIVVH 571
 QY 277 IGFDTVTNNGTHFNPLKKGDEVKAGELCEFDIDAIIKAAGYEVVTPPIVVSNNYKKTGPVN 336
 Db 572 MGDITVALSGKFKRLVEGSDTVKAGEPILEMDLDFLNANARSMISPVVCSNDDISALV 631
 QY 337 TYGLGIEIAG 346
 Db 632 ILASGKVAVG 641

RESULT 12

ID PTSB_BACSU STANDARD; PRT; 460 AA.
 AC P05306;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PTS SYSTEM, SUCROSE-SPECIFIC IIBC COMPONENT (EIIBC-SCR) (SUCROSE-
 DE PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
 DE (EC 2.7.1.69) (EII-SCR).
 GN SACP OR IPA-49D.

RX MEDLINE=91299743; PubMed=1906345;
 RA Fairbrother W.J., Cavanagh J., Dyson H.J., Plamer A.G. III,
 RA Sutrina S.L., Reizer J., Sailer M.H. Jr., Wright P.E.;
 RT "Polypeptide backbone resonance assignments and secondary structure
 of Bacillus subtilis iifig determined by two-dimensional and
 three-dimensional heteronuclear NMR spectroscopy";
 RL Biochemistry 30:6896-6907(1991).
 RN [6]
 RP STRUCTURE BY NMR OF EIAA DOMAIN.
 RX MEDLINE=98254029; PubMed=9593197;
 RA Chen Y., Case D.A., Reizer J., Sailer M.H. Jr., Wright P.E.;
 RT "High-resolution solution structure of Bacillus subtilis IIAglc.";
 RL Proteins 31:258-270(1998).
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IIDC DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.
 CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
 CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
 CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
 CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcements>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z11744; CAA77803.1; -
 CC EMBL; X12832; CAA31315.1; -
 CC EMBL; M60344; AAA22498.1; -
 CC EMBL; Z99111; CAB13262.1; -
 CC PIR; S22752; WQBSGS.
 CC PDB; 1GPR; 31-OCT-93.
 CC PDB; 1AX3; 17-JUN-98.
 CC Subtilist; BG10198; ptsg.
 CC InterPro; IPR001127; PTS_EIIA.
 CC InterPro; IPR001996; PTS_EIIB.
 CC InterPro; IPR003352; PTS_EIIC.
 CC Pfam; PF00358; PTS_EIIA_1; 1.
 CC Pfam; PF00367; PTS_EIIB; 1.
 CC Pfam; PF02378; PTS_EIIC; 1.
 CC ProDom; PD001476; PTS_EIIB; 1.
 CC ProDom; PD002243; PTS_EIIA_1; 1.
 CC ProSITE; PS00371; PTS_EIIA_1; 1.
 CC ProSITE; PS10035; PTS_EIIB_CYS; 1.
 CC Phosphorylation; Transmembrane; Sugar transport; Transferase;
 KW Phosphotransferase system; 3D-structure; Complete proteome.
 FT MOD_RES 239 239 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 461 461 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 620 620 PHOSPHORYLATION (BY SIMILARITY).
 FT DOMAIN 1 ? EIIC DOMAIN.
 FT DOMAIN 1 ? EIIB DOMAIN.
 FT DOMAIN 530 690 EIIA DOMAIN.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 89 109 POTENTIAL.
 FT TRANSMEM 139 159 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT TRANSMEM 283 303 POTENTIAL.
 FT TRANSMEM 314 334 POTENTIAL.
 FT TRANSMEM 338 358 POTENTIAL.
 FT TRANSMEM 365 385 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 FT STRAND 544 544
 FT TURN 546 547

FT STRAND 550 550
 FT STRAND 553 553
 FT STRAND 562 562
 FT HELIX 563 565
 FT HELIX 569 572
 FT TURN 573 574
 FT STRAND 578 584
 FT STRAND 588 590
 FT STRAND 595 600
 FT TURN 602 603
 FT STRAND 606 611
 FT TURN 612 613
 FT STRAND 616 620
 FT STRAND 623 623
 FT HELIX 626 628
 FT TURN 629 632
 FT STRAND 633 635
 FT TURN 639 640
 FT STRAND 642 643
 FT TURN 645 646
 FT STRAND 648 652
 FT HELIX 654 660
 FT STRAND 663 663
 FT STRAND 666 670
 FT TURN 675 676
 FT STRAND 679 681
 FT STRAND 686 687
 FT TURN 689 690
 FT STRAND 692 692
 FT STRAND 695 697
 SQ SEQUENCE 699 AA; 75525 MW; 2A14D3C32EE0A9C5 CRC64;

 Query Match 16.4%; Score 299; DB 1; Length 699;
 Best Local Similarity 24.3%; Pred. No. 5.9e-14;
 Matches 119; Conservative 55; Mismatches 134; Indels 182; Gaps 18;

 QY 7 VGGLLFGLVSPVITGLHQSF-PPIELELFN-----OGGSFIFATASMANIAQG- 55
 DB 221 LAAFIFGVIERSLPIPLGLHHIFSPFWYEFFYSKSAAGEIIRGDDQRIF----MAQIKDGV 276
 QY 56 -----AACLAFFFLAKSEKLKLAGASGVSA-----VLGITPEAIF 91
 DB 277 QLTAGTFMTGKYPFMMFGLPAAALAIYHEAKPQNKLVAGIMGSAALTSFLGITPELEF 336
 QY 92 G-----VNLRLRWPF-----I 103
 DB 337 SFLFVAPVLFALHCLFAGLSFMMVOLLNVKIGMTFSGGLIDYFLFGILPNRTAWLVI 396
 QY 104 GIGTAAI--GGALIAL--FNK-----AVALGAAGFLGVVSDIDAPDMVFLVC 149
 DB 397 GLGLAVIYFGRFAIRKFNLTGPGREDAAEETAAPGKTGEAG---DLPYEIL----- 446
 QY 150 VTFFTAFG-----AAIAYGLVLRNRSIDPD----- 176
 DB 447 ----QAMGDQENIKHLDACITRLRTVNDQKVKDRLKQLGASGVLEVGNIIQAFGPR 502
 QY 177 -----ATAAPVPAGTKRAEAE-----PAEFSNDSTIIQAPLTGEATA 214
 DB 503 SDGLKTMODIIAGRKPRPEKTSAQEEVGQVEEVIAEPLQNEIGEVEFVSPITGEIHP 562
 QY 215 LSSVSDAMFASGKLGSGVAIVPTKQLVSPVSGKIVVAFPPSGHAFVTRKADGNSVDIL 274
 DB 563 ITDVPDOVFSCKMMGDGFALIFSEGVIVSPVRGKTLNVFPPTKHAIGLQS---DGGR-EIL 618
 QY 275 MHIGFDVTNLNCTHENPLKKGDEYKAGELICEFDIDAKAAGYEVVTPPIVVS----- 328
 DB 619 IHFGIDTVSLKGEGETSFVSEGDVRPEGQKLLVDLDVAVKPNVPSLMTPIVFTNLAE 678
 QY 329 --YKKTGPVN 336
 DB 679 VSIKASGSVN 688

```
RESULT 14
PTSB_SALTY
ID PTB_SALTY STANDARD; PRT; 456 AA.
AC P08470;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIIBC COMPONENT (EIIBC-SCR) (SUCROSE-
DE PERMEASE IIIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-SCR).
GN SCRA..
OS Salmonella typhimurium.
OG Plasmid pUR400.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88216186; PubMed=3285123;
RA Ebner R., Lengeler J.W.;
RT "DNA sequence of the gene scrA encoding the sucrose transport protein
RT EnzymeII(Scr) of the phosphotransferase system from enteric bacteria:
RT homology of the EnzymeII(Scr) and EnzymeII(Bgl) proteins.";
RL Mol. Microbiol. 2:9-17(1988).
RN [2]
RP REVISIONS.
RX MEDLINE=94018607; PubMed=8412665;
RA Jahreis K., Lengeler J.W.;
RT "Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid
RT repressor for sucrose and D-fructose specific regulons from enteric
RT bacteria.";
RL Mol. Microbiol. 9:195-209(1993).
RN [3]
RP SEQUENCE OF 1-7 FROM N.A.
RX STRAIN=6153-62;
RC MEDLINE=91100329; PubMed=1846143;
RA Hardesty C., Ferran C., Drenzo J.M.;
RT "Plasmid-mediated sucrose metabolism in Escherichia coli:
RT characterization of scrV, the structural gene for a
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system
RT outer membrane porin.";
RL J. Bacteriol. 173:449-456(1991).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
CC COVALENTLY BOUND EIIC DOMAIN. INSTEAD, EII-SCR-MEDIATED
CC PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
CC GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67750; CAA47973.1; .
DR EMBL; Y00541; CAA68605.1; ALT_SEQ.
DR EMBL; X38416; AAA98418.1; .
DR PIR; S01036; WQEST.
DR HSSP; P05053; IIBA.
DR InterPro; IPR001996; PTS_EIIB.
```

```
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Transferase; Sugar transport; Plasmid;
KW Transmembrane; Inner membrane; Phosphorylation.
FT DOMAIN 1 ? EIIIB DOMAIN.
FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 456 AA; 47864 MW; 98A6F1620AE50885 CRC64;

Query Match 16.3%; Score 296.5; DB 1; Length 456;
Best Local Similarity 40.3%; Pred No. 5.4e-14;
Matches 64; Conservative 29; Mismatches 63; Indels 3; Gaps 1;

QY 5 GPVGLLGLVYSPVITGLHQSFPPIELELFNQ---GGSFIFATASMANIAQGAACLA 61
DB 289 GWLAGLLFGLYSVIVITGIHHSFHAVEAGLLGNPSIGVNFLLPIWAMANVAQGGACLA 348
QY 62 FFLAKSEKILKGLAGSCVSAVLGITEPAIFGVNLRWPFIFIGTAAIGGALLIALEFN 121
DB 349 WFKTKDAKIKAITLPSAFAAMLGITEAAIFGINLRVFKPIAALIGGAAGGAWVSVHY 408
QY 122 AVALGAAGFLGVVSDIDPDVMFMFLVCAVVTFFTAAGAAI 160
DB 409 MTAVGLTAIPGMAIVQASLLNVIIGMVIAGVAFVSL 447

RESULT 15
ID PTAA_ECOLI STANDARD; PRT; 648 AA.
AC P09323;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC COMPONENT (EIIABC-NAG)
DE (N-ACETYLGLUCOSAMINE-PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE
DE ENZYME II, ABC COMPONENT) (EC 2.7.1.69) (EII-NAG).
GN NAGE OR PSTN OR H0679.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88212176; PubMed=3284790;
RA Rogers M.J., Ohgi T., Plumbbridge J., Soell D.;
RT "Nucleotide sequences of the Escherichia coli nage and nagB genes:
RT the structural genes for the N-acetylglucosamine transport protein of
RT the bacterial phosphoenolpyruvate: sugar phosphotransferase system
RT and for glucosamine-6-phosphate deaminase.";
RL Gene 62:197-207(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89050950; PubMed=3056518;
RA Peri K.G., Waygood E.B.;
RT "Sequence of cloned enzyme IIN-acetylglucosamine of the
RT phosphoenolpyruvate:N-acetylglucosamine phosphotransferase system of
RT Escherichia coli.";
RL Biochemistry 27:6054-6061(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RX Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RX Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RX Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:23:46 ; Search time 25.85 Seconds

(without alignments)

1066.738 Million cell updates/sec

Title: US-09-604-231-4

Perfect score: 1821

Sequence: 1 YDFGPGVGLLGLVSPV.....IEAGANLLNVAKEAVPATP 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	711	39.0	664	1 B32243	phosphotransferase
2	706.5	38.8	651	2 S44257	phosphotransferase
3	562.5	30.9	630	2 H83686	PTS system, beta-g
4	550.5	30.2	636	2 C83724	PTS system, beta-g
5	515.5	28.3	632	2 S68599	phosphotransferase
6	515	28.3	609	2 I40406	beta-glucoside per
7	503	27.6	609	2 T47097	hypothetical prote
8	492.5	27.0	631	2 B42603	beta-glucoside-spe
9	450	24.7	625	2 C25977	phosphotransferase
10	448	24.6	636	2 D86807	hypothetical prote
11	446.5	24.5	480	2 S39978	scra protein - Sta
12	416	22.8	479	2 F82432	PTS system, sucros
13	412	22.6	479	2 J00781	sucrose uptake pro
14	350.5	19.2	470	2 H83926	PTS system, trehal
15	330	18.1	470	2 C69725	phosphotransferase
16	314	17.2	495	2 D86766	hypothetical prote
17	313.5	17.2	692	1 S46953	phosphotransferase
18	313	17.2	651	1 S18607	phosphotransferase
19	310	17.0	458	2 H83881	PTS system, sucros
20	309	16.7	460	2 A39938	phosphotransferase
21	299	16.4	699	1 W0BSGS	phosphotransferase
22	295	16.2	648	1 W0EC2N	phosphotransferase
23	294	16.1	648	2 E85567	hypothetical prote
24	293.5	16.1	456	2 S62331	phosphotransferase
25	290	15.9	372	2 I39868	sac operon regulat
26	277.5	15.2	455	1 W0EBST	phosphotransferase
27	276	15.2	675	2 D83755	PTS system, glucos
28	273	15.0	459	2 J00293	levansucrase synth
29	271.5	14.9	173	2 C83839	PTS system, glucos

phosphotransferase
phosphotransferase
phosphotransferase
lactose transport
phosphotransferase
phosphotransferase
PTS system, glucos
PTS system, glucos
PTS system, trehal
PTS system enzyme
phosphotransferase
phosphotransferase
probable phosphotr
raffinose carrier

ALIGNMENTS

RESULT 1

B32243

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus m
C:Species: Streptococcus mutans
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B32243

R:Sato, Y.; Poy, F.; Jacobson, G.R.; Kuramitsu, H.K.

J. Bacteriol. 171, 263-271, 1989

A:Title: Characterization and sequence analysis of the scra gene encoding enzyme II(s

A:Reference number: A32243; MUID:89123027

A:Accession: B32243

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-664 <SAT>

A:Cross-references: GB:M22711; NID:g153799; PIDN:AAA26971.1; PID:g153801

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

C:Keywords: phosphoprotein; phosphotransferase

F:511-664/Domain: phosphotransferase system glucose-specific enzyme II, factor III ho

Query Match 39.0%; Score 711; DB 1; Length 664;

Best Local Similarity 42.1%; Pred. No. 5.4e-43;

Matches 155; Conservative 73; Mismatches 116; Indels 24; Gaps 8;

Qy 1 YDFGPGVGLLGLVSPVITGLHQSPPIEL----FNQG---GSFIFATASMANIAQ 54

Db 307 YDTTGLGCVGEGALYSPVMTGLHQSPPIEL----FNQG---GSFIFATASMANIAQ 366

Qy 55 GAACLAFFLAKSEKLGAGASVAVLGIETPAIFGVNLRWPFIFIGTAAIGGAL 114

Db 367 GAATFAIYFLTKDKMKKGLSSSGVSALLGITPAIFGVNLRWPFIFIGTAAIGGAL 426

Qy 115 IALNKAVALGAAGFLGVSTIDAPVMVFLVCVVTFIAGAAIAYGLVLRNGSID 174

Db 427 AGLLQVAVSLGAGFLGLSLKASSIFPVVCELIISFAIAFAVITYG-----KTKAVD 481

Qy 175 PDATAAPYAGTAKAEAPAEFFSNDST------IIQAPLTGEAIALSSVSDAMFASGK 227

Db 482 VFAAEAEEAEVQ-ETPEEAASAANKAQVTDVLAAPLAGEAVELTSVNDPVFSEA 540

Qy 228 LGSVAIVPTKQLVSPVSGKIVAFPSGHAFVATKEDGSDNDILMHIGFDTVNLNGT 287

Db 541 MGKGIAIRPSGNTVYAPVDGTVOIAFDTCGHAYGI--KSDNGA--BILIHIGIDTVSMGK 596

Qy 288 HFNPLKKGDEKAGELCEFDIDAIAKAGVEVTPPIVVSNNKKTGPVNTY-GLGEIAG 346

Db 597 GFEQKVQADQIKKGDVLGTFDSKDIABAGLDNTTTFIVTNTADYASVETLASSGTVAVG 656

Qy 347 ANLLNVAK 354

Db 657 DSLLEVKK 664

RESULT 2

S44257
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pediococcus pentosae
C:Species: Pediococcus pentosaceus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999
C:Accession: S44257
R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.
submitted to the EMBL Data Library, April 1994
A:Description: The sucrose and raffinose operons of *Pediococcus pentosaceus* PPE1.0.
A:Reference number: S44252

A:Accession: S44257
A:Molecule type: DNA
A:Residues: 1-651 <LEE>
A:Cross-references: EMBL:Z32771; NID:g493728; PIDN:CAA83668.1; PID:g475968
C:Genetics:
A:Gene: scrA

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase
C:Keywords: phosphotransferase
F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol

Query Match 38.8%; Score 706.5; DB 2; Length 651;

Best Local Similarity 42.0%; Pred. No. 1.1e-42;

Matches 154; Conservative 64; Mismatches 126; Indels 23; Gaps 5;

QY 1 YDFGGPVGGLLGLVYSPITVITGLHQSFPPIELF---NOGSGSFATASMANIAQGA 56

DB 300 YNSTGWMGIFGLLYSAIVITGLHQTFFAETQLLANVAKTGGSFIFPVASMANICQGA 359

QY 57 ACLAVFLLAKSEKLKGLAGASVAVLGIPTAIFGVNLRRLRMPFFIGIGTAAIGGALIA 116

DB 360 ATLAIFATKSKOKALTSSAGVSALLGITPAIFGVNLRKMPFFVFAATASGASAFGL 419

QY 117 LFNKAVLGAAGFLGWSIDAPDMVFLVCAVVTFFIAFGAAIAYGLLYVRNGSIDPD 176

DB 420 LFHVLVSAMGPASVIGFISIAKSIAPFMSLAVISVWAFIPFTFY---AKRTLGDGDRD 475

QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP 236

DB 476 QVKSAPTSTV-----INVNDRLISAPVTGASELSKQVNDQVFSIEMKGAAIVP 526

QY 237 TKQLVSPVSGKIVAFPPSGHAFVTRKADGSDNVDILMHIGPDTVNLNGTHFNPLKKQG 296

DB 527 SSDQVVAPAGVITVYDSSHAYGIKTTA---GAEILIHGLDVTNLNGEHTFTNVQKG 582

QY 297 DEYKAGELLCEFDIDAIAKAAGYEVTTPIVSVNKKTPVNTYGLGLEIAGANL--LNVA 354

DB 583 DTVHOGDLGTGFDIAALKAAVNDPTVMLTNTANYANVERLKVTVNQAGEQLVALTAPA 642

QY 355 KEAVPAT 361

DB 643 ASSVAAT 649

RESULT 3

H83686
PTS system, beta-glucoside-specific enzyme II, ABC component BH0296 [imported] - Bacillus
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: H83686

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20263314

A:Accession: H83686
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-630 <STO>
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB04015.1; GSPDB:GN00

A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0296

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 30.9%; Score 562.5; DB 2; Length 630;

Best Local Similarity 34.6%; Pred. No. 1.8e-32;

Matches 124; Conservative 78; Mismatches 143; Indels 13; Gaps 4;

QY 1 YDFGGPVGGLLGLVYSPITVITGLHQSFPPIEL-ELFNOGSGFIFATASMANIAQGAACL 59

DB 282 FTFSPLVLAGAIVAGIQVILVIFGIHGIPIILNLSVRGEDIKAVAPAVESQAGAAL 341

QY 60 AVFFLAKSEKLKGLAGASVAVLGIPTAIFGVNLRRLRMPFFIGIGTAAIGGALIALFN 119

DB 342 GVMRLTKNKKLALAGSTITALFGITEPAVYGVTLPLKKPFTMAVISAAGVGAIVHYG 401

QY 120 IKAVLGAAGFLGVSI---DAPDMVFLVCAVVTFFIAFGAAIAYGLLYVRNGSIDPD 176

DB 402 SVAVAPGAPGLTIPTIFPEDGRGFVAFVIAIISVLA-----AVLTIVYFKDPVDDE 456

QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP 236

DB 457 DTLSNMSGSENEVKREDDKKEPSASEIKSPKAGEVVPVLEVDHVFSSGAMGKGVAVRP 516

QY 237 TKQLVSPVSGKIVAFPPSGHAFVTRKADGSDNVDILMHIGPDTVNLNGTHFNPLKKQG 296

DB 517 KEGRLVAPINGVTSLFETKHAIGITS---DNGTEIFIHVGIDTVOLKGEHFTSFIEQG 572

QY 297 DEYKAGELLCEFDIDAIAKAAGYEVTTPIVSVNKKTPVNTYGLGLEIAGANLNVAK 354

DB 573 DEVAAGDVLLEFDVERITAAGYDVIPTVLITNAKQFSNVOTDKREVTSEDLIHVIK 630

RESULT 4

C83724

PTS system, beta-glucoside-specific enzyme II, ABC component bglp [imported] - Bacill

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

C:Accession: C83724

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a

A:Reference number: A83650; MUID:20263314

A:Accession: C83724

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-636 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04314.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: bglp

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera

Query Match 30.2%; Score 550.5; DB 2; Length 636;

Best Local Similarity 36.7%; Pred. No. 1.3e-31;

Matches 132; Conservative 55; Mismatches 150; Indels 23; Gaps 6;

QY 1 YDFGGPVGGLLGLVYSPITVITGLHQSFPPIEL-ELFNOGSGFIFATASMANIAQGAACL 59

DB 294 YNLSPIIAGAFILGGFQVVFIFGLHWGLIPIAINNLVVOGSDPVLAMVFAASFAQIGAVA 353

QY 60 AVFFLAKSEKLKGLAGASVAVLGIPTAIFGVNLRRLRMPFFIGIGTAAIGGALIALFN 119

DB 354 AWWLKIKQKQVTLSPVAFISGIFGVTEPAIVGVTLPLKRPFIISCIIAAAGVGAIGLFR 413

QY 120 IKAVLGAAGFLGVSI---DAPDMVFLVCAVVTFFIAFGAAIAYGLLYVRNGSID 174

DB 414 SQGYIIGLIGIFIPSLHPADGMDAGFVIVAVVAVF-LGFILTYLGLKSGNASDE 472

QY 175 PDATAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAI 234

DB 473 QTETKHTSTGTGKEEE-----ISSPFGNSVITLSEIKRDEAFSSGALGEGIAI 520

Query Match 18.1%; Score 330; DB 2; Length 470;
Best Local Similarity 41.6%; Pred. No. 4e-16;
Matches 67; Conservative 34; Mismatches 58; Indels 2; Gaps 1;

QY	7	VGGLFGLVYSPVITCLHOSFPPIELEFNQ--GGSFIFATASMANIAOGACIAYFEL	64
Db	307	LGGLLYGGFYSALVITGMHHTFLAVDLQLIGSKLGTFLWPMIALSNIAOGSAAIAMMFI	366
QY	65	AKSEKLGLAGAGSVSAVLGITEPAIFGVNLRWRPFFICIGTAAIGGALIALFNKAVA	124
Db	367	VKDEKOKGLSLTSGISAYLGITEPAIFGVNLRYPFFIIMVSSGLAGMYISSQGVLA	426
QY	125	LGAAGFLGVVSDAPDMVWFLVCANVTFFIAGGAATAYGLY	165
Db	427	VGVGVPGIFSIMSQYWGAFAGMAIVLIVPAGTYAYARF	467

Search completed: March 21, 2002, 16:23:47
Job time: 153 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:27:36 ; Search time 29.76 Seconds
(without alignments)
821.709 Million cell updates/sec

Title:
Perfect score: 1821

Sequence: 1 YDFGPGVGLLGLVSPVIV.....IEAGNLLNVAKKAVPATP 362

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 322656 seqs, 67552660 residues

Total number of hits satisfying chosen parameters: 322656

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/pct_NEW_COMB.pcp:*
2: /cgn2_6/ptodata/1/paa/US05_NEW_COMB.pcp:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pcp:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pcp:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pcp:*
6: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep1:*
7: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pcp:*
8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	542	29.8	627	1	PCT-US02-03987-13467
2	542	29.8	627	6	US-09-815-242-13467
3	542	29.8	627	7	US-10-072-851-13467
4	310.5	17.1	679	1	PCT-US02-03987-5658
5	310.5	17.1	679	6	US-09-815-242-5658
6	310.5	17.1	679	7	US-10-072-851-5658
7	310.5	17.1	681	1	PCT-US02-03987-12270
8	310.5	17.1	681	6	US-09-815-242-12270
9	310.5	17.1	681	7	US-10-072-851-12270
10	295	16.2	648	1	PCT-US02-03987-10089
11	295	16.2	648	6	US-09-815-242-10089
12	295	16.2	648	7	US-10-072-851-10089
13	276.5	15.2	473	6	US-09-897-516-6783
14	276.5	15.2	551	1	PCT-US02-03987-10500
15	276.5	15.2	551	6	US-09-815-242-10500
16	276.5	15.2	551	7	US-10-072-851-10500
17	274	15.0	687	1	PCT-US02-03987-5822
18	274	15.0	687	6	US-09-815-242-5822
19	274	15.0	687	7	US-10-072-851-5822
20	274	15.0	719	1	PCT-US02-03987-12942
21	274	15.0	719	6	US-09-815-242-12942
22	274	15.0	719	7	US-10-072-851-12942
23	247.5	13.6	169	6	US-09-897-516-6596
24	229	12.6	280	1	PCT-US02-03987-12479
25	229	12.6	280	6	US-09-815-242-12479

26	229	12.6	280	7	US-10-072-851-12479	Sequence 12479, A
27	226.5	12.4	842	1	PCT-US02-03987-11950	Sequence 11950, A
28	226.5	12.4	842	6	US-09-815-242-11950	Sequence 11950, A
29	226.5	12.4	842	7	US-10-072-851-11950	Sequence 11950, A
30	221.5	12.2	263	1	PCT-US02-03987-13018	Sequence 13018, A
31	221.5	12.2	263	6	PCT-US02-03987-13152	Sequence 13152, A
32	221.5	12.2	263	7	US-09-815-242-13018	Sequence 13018, A
33	221.5	12.2	263	6	US-09-815-242-13152	Sequence 13152, A
34	221.5	12.2	263	7	US-10-072-851-13018	Sequence 13018, A
35	221.5	12.2	263	6	PCT-US02-03987-5842	Sequence 5842, Ap
36	220.5	12.1	135	1	PCT-US02-03987-5842	Sequence 5842, Ap
37	220.5	12.1	135	6	US-09-815-242-5842	Sequence 5842, Ap
38	220.5	12.1	135	7	US-10-072-851-5842	Sequence 5842, Ap
39	220.5	12.1	484	1	PCT-US02-03987-10809	Sequence 10809, A
40	220.5	12.1	484	6	US-09-815-242-10809	Sequence 10809, A
41	220.5	12.1	484	7	US-10-072-851-10809	Sequence 10809, A
42	201.5	11.1	142	1	PCT-US02-03987-5718	Sequence 5718, Ap
43	201.5	11.1	142	6	US-09-815-242-5718	Sequence 5718, Ap
44	201.5	11.1	142	7	US-10-072-851-5718	Sequence 5718, Ap
45	190.5	10.5	484	1	PCT-US02-03987-12272	Sequence 12272, A

ALIGNMENTS

RESULT 1
PCT-US02-03987-13467
; Sequence 13467, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; TITLE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; PRIOR FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13467
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
PCT-US02-03987-13467

Query Match 29.8%; Score 542; DB 1; Length 627;

Best Local Similarity 36.0%; Pred. No. 1.3e-32;

Matches 128; Conservative 74; Mismatches 130; Indels 24; Gaps 9;

QY	3	FGPGVGLLGLVSPDIVTGLHQSPPIELFNOGGSFIF-ATASMANIAOGAACLAV	61
DB	288	FG-LGGFLIGVGHQILVYSGVHHIENLLEVLQAAADHANPENAIITAAMTAOGAATVAV	345
QY	62	FFLAKSEKLKLAGASGSAVLGITEPAIFGVNLRWRPFIFIGITAAIGGALIALFNK	121
DB	346	GVKTNPKLTKLAFPAALSAFLGITEPAIFGVNLRWRPFIFIGITAAIGGALIALFNK	405
QY	122	AVALGAAGFLG-VVSDADPMVFLCAVVTFFIAFGAAIAYGLVLRNRSIDPDATAA	180
DB	406	GTGNGITITPTMLYVGNQLPQYLLMVAVSFALGALTVMFCY-----EDEVATAA	458
QY	181	PYPAGTKRAAE-APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKG	239
DB	459	AKQAEVAEKEEVAAPALQNETLV--TPIVGDVVALADVNDPVFSSGAMGQGIKVPKSG	516
QY	240	QLVSPVSGKIVAVFSGHFAVTRKEDSNVDILMHIGFDIVNLGHTFNPLKKGQDEV	299
DB	517	VYVALADAEVSAFPTGHFGLKTR---NGAEVLHVIGDITVSMNGDGFKAQVGNKV	572
QY	300	KAGELLCEDIDAIKAAGVEVTTPIVWSN--YKKTGPVNTYGLGIEAGANLLN	352
DB	573	KAGDVLGTFDSNKIAAAGLUDDITMIVTNTADYASVAPVAT---GSVSKGDAVIEV	625

RESULT 2

US-09-815-242-13467

; Sequence 13467, Application US/09815242

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13467

; LENGTH: 627

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-815-242-13467

Query Match

Best Local Similarity 29.8%; Score 542; DB 6; Length 627;

Matches 128; Conservative 74; Mismatches 130; Indels 24; Gaps 9;

QY 3 FGGPVGGLFGLVYSPVITGLHQSPFPIELELFNQGGSFIF-ATASMANIAQGAACLV 61

Db 288 FG--LGGFLIGGVHQLIVVSGVHHIFNLLEVLQLLAADHANPFAITTAAMTAQGAATVAV 345

QY 62 FFLAKSEKLLKLAGAGSCVAVLGITPAIFGVNLRWPPFFIGTAAIGGALIALFNK 121

Db 346 GVTKNPKLKTALFAPPAALSALFAGITPAIFGVNLRWPPFFLSLIAGIAGGGLASILGLA 405

QY 122 AVALGAAGFLG-VVSDAPDMVLFCAVVTFFIAFGAAIAYGLYLVRNGSIDPDATAA 180

Db 406 GTGNGITIPGTMLYVGVNGOLPOYLLMVAVSFALGFALTVMFCY-----EDEVATAA 458

QY 181 PVPAGTTKAEAE-APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKG 239

Db 459 AKQAEVAEEKEEVAPAAALQNETLV--TPIVGDVVALADVNDPVFSSGAMGGIAPKPSQG 516

QY 240 QLVSPVSGKIVAFPPSGHAFVTRKADGNSVDILMHIGFTVNLNTHNPLKKGDEV 299

Db 517 VVYALADAEVSIAPFTGHAFGLKTR----NGAEVLHIVGIDTVSMNGDGFEEKVAQGNKV 572

QY 300 KAGELICEFDIDAIAKAGYEVVTPPIVVSN---YKKTGPVNTYGLGTEAGANLLNV 352

Db 573 KAGDVLGTGDSNKKIAAAGLDDTTWIVTNTADYASVAPVAT---GSVSKGDVAVIEV 625

RESULT 3

US-10-072-851-13467

; Sequence 13467, Application US/10072851

; GENERAL INFORMATION:

; APPLICANT: Carr, Grant J.

; APPLICANT: Xu, H. Howard

; APPLICANT: Foulkes, J. Gordon

; APPLICANT: Zamudio, Carlos

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Roemer, Terry

; APPLICANT: Jiang, Bo

; APPLICANT: Boosey, Charles

; APPLICANT: Bussey, Howard

; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit

; FILE OF INVENTION: Proliferation

; FILE REFERENCE: ELITRA.028A

; CURRENT APPLICATION NUMBER: US/10/072,851

; CURRENT FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; NUMBER OF SEQ ID NOS: 15811

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13467

; LENGTH: 627

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-10-072-851-13467

Query Match

Best Local Similarity 29.8%; Score 542; DB 7; Length 627;

Matches 128; Conservative 74; Mismatches 130; Indels 24; Gaps 9;

QY 3 FGGPVGGLFGLVYSPVITGLHQSPFPIELELFNQGGSFIF-ATASMANIAQGAACLV 61

Db 288 FG--LGGFLIGGVHQLIVVSGVHHIFNLLEVLQLLAADHANPFAITTAAMTAQGAATVAV 345

QY 62 FFLAKSEKLLKLAGAGSCVAVLGITPAIFGVNLRWPPFFIGTAAIGGALIALFNK 121

Db 346 GVTKNPKLKTALFAPPAALSALFAGITPAIFGVNLRWPPFFLSLIAGIAGGGLASILGLA 405

QY 122 AVALGAAGFLG-VVSDAPDMVLFCAVVTFFIAFGAAIAYGLYLVRNGSIDPDATAA 180

Db 406 GTGNGITIPGTMLYVGVNGOLPOYLLMVAVSFALGFALTVMFCY-----EDEVATAA 458

QY 181 PVPAGTTKAEAE-APAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKG 239

Db 459 AKQAEVAEEKEEVAPAAALQNETLV--TPIVGDVVALADVNDPVFSSGAMGGIAPKPSQG 516

QY 240 QLVSPVSGKIVAFPPSGHAFVTRKADGNSVDILMHIGFTVNLNTHNPLKKGDEV 299

Db 517 VVYALADAEVSIAPFTGHAFGLKTR----NGAEVLHIVGIDTVSMNGDGFEEKVAQGNKV 572

QY 300 KAGELICEFDIDAIAKAGYEVVTPPIVVSN---YKKTGPVNTYGLGTEAGANLLNV 352

Db 573 KAGDVLGTGDSNKKIAAAGLDDTTWIVTNTADYASVAPVAT---GSVSKGDVAVIEV 625

RESULT 4

PCT-US02-03987-5658

; Sequence 5558, Application PC/TUS0203987

; GENERAL INFORMATION:

; APPLICANT: Elittra Pharmaceuticals, Inc.

; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit

; FILE OF INVENTION: Proliferation

; FILE REFERENCE: ELITRA.028VPC

; CURRENT APPLICATION NUMBER: PCT/US02/03987

; CURRENT FILING DATE: 2002-02-02

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
PCT-US02-03987-5658

Query Match 17.1%; Score 310.5; DB 1; Length 679;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVYSPVITGLHQSF-----PPIELELFNOG----- 39
DB 209 VAVFLFGFKRLILPGLHHIFHAPFWFEGSNKNAAGIIGHGQRIEIQREGAHLTA 268
QY 40 GSFIATASMANIAQGAACLAFLAKSEKLGAGASGVA-----VLGITEPAIFGVNL 95
DB 269 GKFMQGEFPVMFGLPAAALAIYHSAPENKVVAGLMGSAALTSFLTGTITEPLEFSFLF 328
QY 96 RLRFWFFIGITAAAGGALIALFNKAVLG---AAGFLGVVSIDA-PDMVMF----- 144
DB 329 VAPLLEFFI---HAVLDGLSFLTLYLDDVHLGYTFSGGFDYVLLGVLPNKTQMWLVIPVG 385
QY 145 LVCVVTFPI-----AFGAIAAYGLY----- 165
DB 386 LVAVIYFVFRFLIVKLKYKTPGREDKQSOAVTASATELPYAVLEAMGKANIKHLDAC 445
QY 166 LVR-----RNGSIDPD 176
DB 446 ITRLRVEVNDKSKVDVPGKDLGASGVLEVGNMMAIFGPKSDQIKHEMQQIMNGQVVEN 505
QY 177 ATAAPVPAGTTKAEAPAEFNSDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP 236
DB 506 PTTMEDDKDQETVVVAEDKSAATSELSHIVHAPLTGEVTPLEVPDQVFSEKMMGDGIAIKP 565
QY 237 TKGOLVSPSGKIVVAFPSGHAFVTRKAEDGNSVDILMHIGFDTVNLNGTHFNPLKKOG 296
DB 566 SQGEVRAPNGKVMIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 621
QY 297 DEVKAGELCEFDIDAIAKAAGYEVTTPIVVS 327
DB 622 QEVKQGDLLINFDLDYIRNHAKS DITPIIYT 652

RESULT 5
US-09-815-242-5658
; Sequence 5658, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5658
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5658

Query Match 17.1%; Score 310.5; DB 6; Length 679;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;
QY 7 VGGLLFGLVYSPVITGLHQSF-----PPIELELFNOG----- 39
DB 209 VAVFLFGFKRLILPGLHHIFHAPFWFEGSNKNAAGIIGHGQRIEIQREGAHLTA 268
QY 40 GSFIATASMANIAQGAACLAFLAKSEKLGAGASGVA-----VLGITEPAIFGVNL 95
DB 269 GKFMQGEFPVMFGLPAAALAIYHSAPENKVVAGLMGSAALTSFLTGTITEPLEFSFLF 328
QY 96 RLRFWFFIGITAAAGGALIALFNKAVLG---AAGFLGVVSIDA-PDMVMF----- 144
DB 329 VAPLLEFFI---HAVLDGLSFLTLYLDDVHLGYTFSGGFDYVLLGVLPNKTQMWLVIPVG 385
QY 145 LVCVVTFPI-----AFGAIAAYGLY----- 165
DB 386 LVAVIYFVFRFLIVKLKYKTPGREDKQSOAVTASATELPYAVLEAMGKANIKHLDAC 445
QY 166 LVR-----RNGSIDPD 176
DB 446 ITRLRVEVNDKSKVDVPGKDLGASGVLEVGNMMAIFGPKSDQIKHEMQQIMNGQVVEN 505
QY 177 ATAAPVPAGTTKAEAPAEFNSDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVP 236
DB 506 PTTMEDDKDQETVVVAEDKSAATSELSHIVHAPLTGEVTPLEVPDQVFSEKMMGDGIAIKP 565
QY 237 TKGOLVSPSGKIVVAFPSGHAFVTRKAEDGNSVDILMHIGFDTVNLNGTHFNPLKKOG 296
DB 566 SQGEVRAPNGKVMIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 621
QY 297 DEVKAGELCEFDIDAIAKAAGYEVTTPIVVS 327
DB 622 QEVKQGDLLINFDLDYIRNHAKS DITPIIYT 652

RESULT 6
US-10-072-851-5658
; Sequence 5658, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072,851
; CURRENT FILING DATE: 2002-02-08

;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; NUMBER OF SEQ ID NOS: 15811
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5658
;; LENGTH: 679
;; TYPE: PRT
;; ORGANISM: Staphylococcus aureus
US-10-072-851-5658

Query Match 17.1%; Score 310.5; DB 7; Length 679;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

```
QY 7 VGLLFLGLVSPVITGLHOSF-----PPIELELFNOG----- 39
Db 209 VAVFLGFKRLLPGLHIFHAPFWFEGSKWNAAGEIHHGDQRIEIQREGAHLTA 268
QY 40 GSFIATASMANIAQAACLAFLAKSEKLGLAGASGVA-----VLGITEPAIFGVNL 95
Db 269 GKFMQGEFFVMFGLPAAALAIYHSAPENKPKVYVAGLMSAALTSTLTGITEPLEFSFLF 328
QY 96 RLKWPFFIGTGAAGGALIALFNKAVALG---AAGFLGVVSDA--PDMVMF----- 144
Db 329 VAPLLFFI---HAVLDGLSFLTLVLLDVHLGYTFSGGFIDYVLLGVLPNKQTQWMLVIPVG 385
QY 145 LVCVAVTFFI-----AFCAAIAYGLY----- 165
Db 386 LVAVIYVFFRFLVVKLYKTPGREDKQSOAVTASATELPYAVLEAMGKANIKHLDAC 445
QY 166 LVR-----RNGSIDDP 176
Db 446 ITRLRVEVNDKSKVDVPGLKDLGASGVLEGVNNMQAIFGPKSDQIKHEMQOIMNGQVVEN 505
QY 177 ATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVSDAMPASGKLGSGVAIVP 236
Db 506 PTMEDDKDETIVVVAEDKTSATSELSHIVHAPLTGEVTPLEVPDQVSEKMMGSDGIAIKP 565
QY 237 TKGOLVSPVSGKIVVAFPPSGHAFVTRTKAEDGSNVDILMHIGFDVTNLTGTHFNPLKKQG 296
Db 566 SQGEVRAPFNGKVMIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 621
QY 297 DEVKAGELLCFDDAIIKAAGYEVTTPIVVS 327
Db 622 QEVKQGDLLNFDLDYIRNHAQSDITPIIVT 652
```

RESULT 7
PCT-US02-03987-12270
; Sequence 12270, Application PCT/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for identifying the Target of a Compound which Inhibits
; FILE OF INVENTION: Proliferation
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12270
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
PCT-US02-03987-12270

Query Match 17.1%; Score 310.5; DB 1; Length 681;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

```
QY 7 VGLLFLGLVSPVITGLHOSF-----PPIELELFNOG----- 39
Db 211 VAVFLGFKRLLPGLHIFHAPFWFEGSKWNAAGEIHHGDQRIEIQREGAHLTA 270
QY 40 GSFIATASMANIAQAACLAFLAKSEKLGLAGASGVA-----VLGITEPAIFGVNL 95
Db 271 GKFMQGEFFVMFGLPAAALAIYHSAPENKPKVYVAGLMSAALTSTLTGITEPLEFSFLF 330
QY 96 RLKWPFFIGTGAAGGALIALFNKAVALG---AAGFLGVVSDA--PDMVMF----- 144
Db 331 VAPLLFFI---HAVLDGLSFLTLVLLDVHLGYTFSGGFIDYVLLGVLPNKQTQWMLVIPVG 387
QY 145 LVCVAVTFFI-----AFCAAIAYGLY----- 165
Db 388 LVAVIYVFFRFLVVKLYKTPGREDKQSOAVTASATELPYAVLEAMGKANIKHLDAC 447
QY 166 LVR-----RNGSIDDP 176
Db 448 ITRLRVEVNDKSKVDVPGLKDLGASGVLEGVNNMQAIFGPKSDQIKHEMQOIMNGQVVEN 507
QY 177 ATAAPVPAGTTKAEAPAEFSDNSTIIQAPLTGEATLSSVSDAMPASGKLGSGVAIVP 236
Db 508 PTMEDDKDETIVVVAEDKTSATSELSHIVHAPLTGEVTPLEVPDQVSEKMMGSDGIAIKP 567
QY 237 TKGOLVSPVSGKIVVAFPPSGHAFVTRTKAEDGSNVDILMHIGFDVTNLTGTHFNPLKKQG 296
Db 568 SQGEVRAPFNGKVMIFPTKHAIGL-----VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 623
QY 297 DEVKAGELLCFDDAIIKAAGYEVTTPIVVS 327
Db 624 QEVKQGDLLNFDLDYIRNHAQSDITPIIVT 654
```

RESULT 8
US-09-815-242-12270
; Sequence 12270, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12270
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12270

Query Match 17.1%; Score 310.5; DB 6; Length 681;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;
Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVYSPVITGLHQSFP-----PPLELFNQ----- 39
DB 211 VAVLFGLFKRLLIPFGLHHIFHAPFWEFGSWKNAAGEIIHQDRIETIEQIREGAHLTA 270
QY 40 GSFIFATASMANIAOAGACLAFFELAKSEKLAGAGSVSA-----VLGITPAIFGVNL 95
DB 271 GKFMGEPFVMMFGLPAAALAIYHSAPKPKNVVAGLMSAALTSTLTGITEPLEFSELF 330
QY 96 RLRPWFPGIGTAAGGALIALFNKAVAGL---AAGFLGVVSDA---PDMVNF----- 144
DB 331 VAPLLFFI---HAVLDGLSFLTYLLDVLHGYTSGGFDIVLLGVLPNKQWLPIVPG 387
QY 145 LVCVVTFPI-----AFGAAIAYGLY----- 165
DB 388 LVYAVIYVVFRELIVKLKYPGREDKQSAVTASATELPYAVLEAMGGRKANIKHLDAC 447
QY 166 LVR-----RNGSIDPD 176
DB 448 ITRLREVNDSKVDVPGKLDGASGVLEVGNMNAIFGPKSDQIKHEMQQIMNGOVVEN 507
QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGAEIALSSVSDAMFASGKLGSGVAIYP 236
DB 508 PTTMEDDKDETIVVNAEDKSATSELSHIVHAPLTGEVTPLEVPDQVFESEKMMGDGIAIKP 567
QY 237 TKGOLVSPVSGKIVVAPPSGHAFVTRKAEDGSNDVILMHIGFDIVNLNGTHFNPLKKQG 296
DB 568 SOGEVRAPFNGKVMIFPTKHAIGL---VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 623
QY 297 DEVKAGELLCFEDDAIKAAAGVETVPIVVS 327
DB 624 QEVKQGDLLINFEDLDYIRNHAKSIDTPIIYT 654

RESULT 9
US-10-072-851-12270
; Sequence 12270, Application US/10072851
; GENERAL INFORMATION:
; APPLICANT: Carr, Grant J.
; APPLICANT: Xu, H. Howard
; APPLICANT: Foulkes, J. Gordon
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Roemer, Terry
; APPLICANT: Jiang, Bo
; APPLICANT: Boone, Charles
; APPLICANT: Bussey, Howard
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibits C
; FILE REFERENCE: ELITRA.028A
; CURRENT APPLICATION NUMBER: US/10/072.851
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12270
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-072-851-12270

Query Match 17.1%; Score 310.5; DB 7; Length 681;
Best Local Similarity 24.6%; Pred. No. 2.6e-15;

Matches 111; Conservative 54; Mismatches 149; Indels 137; Gaps 11;

QY 7 VGGLLFGLVYSPVITGLHQSFP-----PPLELFNQ----- 39
DB 211 VAVLFGLFKRLLIPFGLHHIFHAPFWEFGSWKNAAGEIIHQDRIETIEQIREGAHLTA 270
QY 40 GSFIFATASMANIAOAGACLAFFELAKSEKLAGAGSVSA-----VLGITPAIFGVNL 95
DB 271 GKFMGEPFVMMFGLPAAALAIYHSAPKPKNVVAGLMSAALTSTLTGITEPLEFSELF 330
QY 96 RLRPWFPGIGTAAGGALIALFNKAVAGL---AAGFLGVVSDA---PDMVNF----- 144
DB 331 VAPLLFFI---HAVLDGLSFLTYLLDVLHGYTSGGFDIVLLGVLPNKQWLPIVPG 387
QY 145 LVCVVTFPI-----AFGAAIAYGLY----- 165
DB 388 LVYAVIYVVFRELIVKLKYPGREDKQSAVTASATELPYAVLEAMGGRKANIKHLDAC 447
QY 166 LVR-----RNGSIDPD 176
DB 448 ITRLREVNDSKVDVPGKLDGASGVLEVGNMNAIFGPKSDQIKHEMQQIMNGOVVEN 507
QY 177 ATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGAEIALSSVSDAMFASGKLGSGVAIYP 236
DB 508 PTTMEDDKDETIVVNAEDKSATSELSHIVHAPLTGEVTPLEVPDQVFESEKMMGDGIAIKP 567
QY 237 TKGOLVSPVSGKIVVAPPSGHAFVTRKAEDGSNDVILMHIGFDIVNLNGTHFNPLKKQG 296
DB 568 SOGEVRAPFNGKVMIFPTKHAIGL---VSDSGLELLIHIGLDTVKLNGEGFTLHVEEG 623
QY 297 DEVKAGELLCFEDDAIKAAAGVETVPIVVS 327
DB 624 QEVKQGDLLINFEDLDYIRNHAKSIDTPIIYT 654

RESULT 10
PCT-US02-03987-10089
; Sequence 10089, Application PC/TUS0203987
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals, Inc.
; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit
; FILE REFERENCE: ELITRA.028VPC
; CURRENT APPLICATION NUMBER: PCT/US02/03987
; CURRENT FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 15811
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10089
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Escherichia coli
PCT-US02-03987-10089

Query Match 16.2%; Score 295; DB 1; Length 648;
Best Local Similarity 23.2%; Pred. No. 3.5e-14;
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;

QY 5 GPVGGLLFGLVYSPVITGLHQSFPPI-----ELELFNQ-----GSF 42
DB 170 GALSGIFGINRLIPTGLHQLVNTIAWFQIGETNAAGTVFPHGDINIRAGDGTAGMF 229
QY 43 IPATASMANIAOAGACLAFFELAKSEK---LKGAGASGVSAVL-GITPEPAIF----- 91
DB 230 MSGFFPIMMFGLPGAALAMYFAAPKRRPVMVGMLLSVAVTAFLTGTGTEPLELFMLAP 289
QY 92 -----GVNRLRWPFFIGTGTAAIGGAL----- 114
DB 290 LLYLLHALLTGISLGFATLGIHAGFSFSAIDYALMNLPAASQNVNMLLMGVIFFA 349
QY 115 -----IALFNKA-----VALGAA----- 128

Db 350 IYFVFSLVIRNFKTPGREDKEDEIVTEANSNTEGLTQLATNYIAAVGTDNLKAI 409
QY 129 -----GFLGVVSIDAPDMVNLVCAVVTFFIAFGAAIAYG 163
Db 410 DACITRLRLTVADSRVNDTMCKRLGASGVVKNKOTIQIVGAKAESIGDAMKKVVARG 469
QY 164 LYLVRNGSIDPATAAPVPAGTTKAEAPAEFSNDSTIIIOAPLTGEAIALSSVSDAMF 223
Db 470 -PVAASAEATP-ATAAPVAKPOAVPNVSAIE-----LVSPITGDDVVALDOVPDEAF 520
QY 224 ASKGLSGVAIVPTKQGLVSPVSGKIVVAPPSGHAFVARTKAEAGSNDVILMHIGFDTVN 283
Db 521 ASKAVGDGVAVKPTDKIVVSPAAGTIVKIFNTNHAFCLETE-----KGAEIVVHMGIDTVA 576
QY 284 LNGTHENPLKKGDEYKAGELCEFDIDAIAKAAGYEVTTPVIVSNYKKTGPVNTYGLGEI 343
Db 577 LEGKGFRLVEGAQVSAGQPILEMDLDVNLNANARSMISPVVCSNIDDFSGLIIRKAGGHI 636
QY 344 EAGANLLNVAKK 355
Db 637 VAGQTPLYEIKK 648

RESULT 11

US-09-815-242-10089

; Sequence 10089, Application US/09815242

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows version 4.0

; SEQ ID NO 10089

; LENGTH: 648

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-815-242-10089

Query Match 16.2%; Score 295; DB 6; Length 648;
Best Local Similarity 23.2%; Pred. No. 3.5e-14;
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;

QY 5 GPVGGLLGLVSPVITGLHQSFPPI-----EELFNQG----GSF 42
Db 170 GALSIGIFGINRLIPTGLHOVLNTIAWFQIGFTNAAGTVPHGDIINRFYAGDGTAGMF 229
QY 43 IPATASMANIAOGAAACLAFFVFLAKSEK---LKGLAGAGVSAVL-GITEPAIF----- 91

Db 230 MSGFFPIMMFGPLPGAALAMYFAAPKERRPMVGGMLLSVAVTAFLTGTVEPLEFLMFAP 289
QY 92 -----GVNLRLRWPFFIGTAAIGGAL----- 114
Db 290 LLYLLHALITGISLFTVATLLGIHAGFSFSAGADYALMYNLPAASONVMMLLVMGVIFFA 349
QY 115 -----IALFNKA-----VALGAA----- 128
Db 350 IYFVFSLVIRNFKTPGREDKEDEIVTEANSNTEGLTQLATNYIAAVGTDNLKAI 409
QY 129 -----GFLGVVSIDAPDMVNLVCAVVTFFIAFGAAIAYG 163
Db 410 DACITRLRLTVADSRVNDTMCKRLGASGVVKNKOTIQIVGAKAESIGDAMKKVVARG 469
QY 164 LYLVRNGSIDPATAAPVPAGTTKAEAPAEFSNDSTIIIOAPLTGEAIALSSVSDAMF 223
Db 470 -PVAASAEATP-ATAAPVAKPOAVPNVSAIE-----LVSPITGDDVVALDOVPDEAF 520
QY 224 ASKGLSGVAIVPTKQGLVSPVSGKIVVAPPSGHAFVARTKAEAGSNDVILMHIGFDTVN 283
Db 521 ASKAVGDGVAVKPTDKIVVSPAAGTIVKIFNTNHAFCLETE-----KGAEIVVHMGIDTVA 576
QY 284 LNGTHENPLKKGDEYKAGELCEFDIDAIAKAAGYEVTTPVIVSNYKKTGPVNTYGLGEI 343
Db 577 LEGKGFRLVEGAQVSAGQPILEMDLDVNLNANARSMISPVVCSNIDDFSGLIIRKAGGHI 636
QY 344 EAGANLLNVAKK 355
Db 637 VAGQTPLYEIKK 648

RESULT 12

US-10-072-851-10089

; Sequence 10089, Application US/10072851

; GENERAL INFORMATION:

; APPLICANT: Carr, Grant J.

; APPLICANT: Xu, H. Howard

; APPLICANT: Foulkes, J. Gordon

; APPLICANT: Zamudio, Carlos

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Roemer, Terry

; APPLICANT: Jiang, Bo

; APPLICANT: Boone, Charles

; APPLICANT: Bussey, Howard

; TITLE OF INVENTION: Methods for Identifying the Target of a Compound which Inhibit

; FILE REFERENCE: ELITRA.028A

; CURRENT APPLICATION NUMBER: US/10/072,851

; CURRENT FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; NUMBER OF SEQ ID NOS: 15811

; SOFTWARE: FastSeq for Windows version 4.0

; SEQ ID NO 10089

; LENGTH: 648

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-072-851-10089

Query Match 16.2%; Score 295; DB 7; Length 648;
Best Local Similarity 23.2%; Pred. No. 3.5e-14;
Matches 114; Conservative 62; Mismatches 162; Indels 154; Gaps 12;

QY 5 GPVGGLLGLVSPVITGLHQSFPPI-----EELFNQG----GSF 42
Db 170 GALSIGIFGINRLIPTGLHOVLNTIAWFQIGFTNAAGTVPHGDIINRFYAGDGTAGMF 229

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10500
LENGTH: 551
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-815-242-10500

Query Match 15.2%; Score 276.5; DB 6; Length 551;
Best Local Similarity 23.0%; Pred. No. 6.8e-13;
Matches 107; Conservative 68; Mismatches 130; Indels 161; Gaps 17;
QY 5 GPGVGLLFGVLVSPVITGLHQSPPI---ELELFNOGGSFI----- 43
Db 77 GFVGAGLYGFNNLLIPTGLHNLNSVWFDFVAGINDIGNFLAGQQALDTGKAIVGQTGM 136
QY 44 ---FATASMANIAQGAACLAFLAKSEKLGKLAG---ASGVSAVL-GITEPAIFGVNL 95
Db 137 YQAGFFPVMFGLPAGA--PAIQOCARPEKKVKTASIMLAAGFAAFTGVTEPLEFSF-M 193
QY 96 RLKWPFFIGTAAIGGALIALNIKVALG-----AAGFLG-VVSDAPD--- 140
Db 194 FVAMPLVY-----LHAVETGISLAPAEPMHWNTAGFAFSAGFVDFELSLKNPVANH 243
QY 141 ---MYMFLVCVVTFE-----IARF-----AAIAYGLY- 165
Db 244 PMMLVQGLVFAAIYFGFRFAITKFNLMTPGREEGDEETPDVAGDNKFASLARIYD 303
QY 166 -----LVRNGSIDP----- 175
Db 304 GLGADANVTSIDNCTTRLRLTVKDTCKVDQAKIKATGVPGVKVIDDTNIQIVIGTEVQFV 363
QY 176 -----DATAAPV---PAGTTKAEAPAEFNSDSTIIQAPLTGEAIALSSVSAM 222
Db 364 ADEMQRLYNHQAPATPVKETPVSQPVVEKKAPV--STKETELYSVANGKVIPISEVPDDV 421
QY 223 FASGKLGSQVAIVPTKGOLVSPVSGKIVVAFPSGSHAFVARTKAEDGSNDVILMHIGFDTV 282
Db 422 FSAKMGDGFVAVPTDGVSTPVAGKITSTIFPTKHALGIQT-----DSGIEVLHLMGLDTV 477
QY 283 NLNGTHFNPLKQGVKAGELCEFDIDAIAAGYEVTTPIVVSN 328
Db 478 ELOGGPFTHVEESQVVKOGDKIATIDLAAEQAGKKSDLIIVFTN 523

Search completed: March 21, 2002, 16:27:37
Job time: 338 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:23:16 ; Search time 21.58 Seconds
(without alignments)
377.488 Million cell updates/sec

Title: US-09-604-231-4
Perfect score: 1821
Sequence: 1 YDFGPPVGGLLFGLVSPV.....IERGANLLNVAKREAVPATP 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113.5	6.2	1176	1	US-07-828-788A-12
2	113.5	6.2	1176	1	Sequence 12, Appl
3	113.5	6.2	1176	3	Sequence 2, Appl1
4	113.5	6.2	1176	5	Sequence 2, Appl1
5	113.5	6.2	1179	1	Sequence 12, Appl
6	106.5	5.8	1129	6	Sequence 1, Appl1
7	106.5	5.8	1179	6	Patent No. 5164180
8	99.5	5.5	522	4	Patent No. 5188960
9	99.5	5.5	654	4	Sequence 3, Appl1
10	98	5.4	1184	6	Sequence 35, Appl
11	98	5.4	1188	6	Patent No. 5254799
12	97.5	5.4	2005	3	Patent No. 5254799
13	96.5	5.3	593	6	Sequence 7, Appl1
14	96.5	5.3	648	3	Patent No. 5523211
15	96.5	5.3	934	1	Sequence 13, Appl
16	96.5	5.3	934	1	Sequence 7, Appl1
17	96.5	5.3	1011	3	Sequence 2, Appl1
18	96.5	5.3	1176	1	Sequence 6, Appl1
19	96.5	5.3	1176	1	Sequence 10, Appl
20	96.5	5.3	1176	1	Sequence 6, Appl1
21	96.5	5.3	1176	5	Sequence 10, Appl
22	96.5	5.3	1176	5	Sequence 10, Appl
23	96.5	5.3	1984	3	Sequence 10, Appl
24	96.5	5.3	1989	3	Sequence 12, Appl
25	95	5.2	680	2	Sequence 10, Appl
26	94	5.2	1835	3	Sequence 2, Appl1
27	94	5.2	2259	4	Sequence 15, Appl
					Sequence 70, Appl

28 94 5.2 2439 3 US-09-335-409-7 Sequence 7, Appl1
29 92.5 5.1 614 3 US-08-622-740-8 Sequence 8, Appl1
30 92.5 5.1 614 3 US-08-440-689-8 Sequence 8, Appl1
31 92.5 5.1 615 3 US-09-042-426-10 Sequence 10, Appl
32 92.5 5.1 615 4 US-09-291-238-10 Sequence 10, Appl
33 92.5 5.1 615 4 US-09-330-760-10 Sequence 10, Appl
34 92.5 5.1 615 4 US-09-328-473-10 Sequence 10, Appl
35 92.5 5.1 615 4 US-09-330-737-10 Sequence 10, Appl
36 92.5 5.1 695 1 US-07-671-817A-5 Sequence 5, Appl1
37 92.5 5.1 969 1 US-07-671-817A-6 Sequence 6, Appl1
38 92.5 5.1 1155 1 US-08-349-867-19 Sequence 19, Appl
39 92.5 5.1 1155 1 US-08-349-867-33 Sequence 33, Appl
40 92.5 5.1 1155 1 US-08-239-476-19 Sequence 19, Appl
41 92.5 5.1 1155 1 US-08-446-486-2 Sequence 2, Appl1
42 92.5 5.1 1155 1 US-07-951-715A-9 Sequence 9, Appl1
43 92.5 5.1 1155 1 US-08-463-308-2 Sequence 2, Appl1
44 92.5 5.1 1155 2 US-08-598-305A-19 Sequence 19, Appl
45 92.5 5.1 1155 2 US-08-598-305A-33 Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-07-828-788A-12
; Sequence 12, Application US/07828788A
; Patent No. 5273746
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: HICKLE, LESLIE A.
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/828,788A
; FILING DATE: 19920129
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MAY5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1176 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: AIZAWAI
; INDIVIDUAL ISOLATE: PS811
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 811A2
; US-07-828-788A-12

GENERAL INFORMATION:

APPLICANT: Payne, Jewel M
APPLICANT: Sick, August J
TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 54078
TITLE OF INVENTION: Lepidopteran-active Toxins
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st STREET, SUITE A-1
CITY: GAINESVILLE
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/040,751
FILING DATE: 19930329
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REFERENCE/DOCKET NUMBER: MA39.C1.D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1179 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: AIZAWAI
INDIVIDUAL ISOLATE: PS811
IMMEDIATE SOURCE:
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
CLONE: 81LR1
US-08-040-751-1

Query Match 6.2%; Score 113.5; DB 1; Length 1179;
Best Local Similarity 22.4%; Pred. No. 0.0076;
Matches 56; Conservative 43; Mismatches 96; Indels 55; Gaps 13;

QY 143 MFLVCVVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 188
Db 380 LFVLDGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPDQNSVVPARAGFSHRLSHVTML 437
QY 189 AEA-----EAP-----AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 230
Db 438 SQAAGAVYTLRATFTSWRHSAEFSLIPSSQITQIPLT-KSINLGSSTSVVKGPGFTGG 496
QY 231 GVAIVPTKGQVSPVSGKIVVAFPSGHAFVRTKAEDGSNVVDILMHIGFTVNLNGTHFN 290
Db 497 DILRITSPGQISTL---RVITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551
QY 291 PLKKQDEVKAGELLCEDFIDAIAKAGYEVVTPIVVSNYKKTGPVNYGLGEIAGANLL 350
Db 552 ATMSSGGNLQSG-----SFRTAGF--TTPF---NFSNGSSIETLSAHVFNNSG-NEV 596
QY 351 NVAKKEAVPA 360
Db 597 YIERIEFVPA 606

RESULT 6

5164180-6
Patent No. 5164180
APPLICANT: Payne, Jewel; Sick, August J.
TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE
AGAINST LEPIDOPTERAN PESTS
NUMBER OF SEQUENCES: 6
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/451,389
FILING DATE: 14-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 353,860
FILING DATE: 18-MAY-1989
SEQ ID NO: 6:
LENGTH: 1129
5164180-6

Query Match 5.8%; Score 106.5; DB 6; Length 1129;
Best Local Similarity 22.4%; Pred. No. 0.039;
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;

QY 143 MFLVCVVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 188
Db 380 LFVLDGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPDQNSVVPARAGFSHRLSHVTML 437
QY 189 AEA-----EAP-----AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 230
Db 438 SQAAGAVYTLRATFTSWRHSAEFSLIPSSQITQIPLT-KSINLGSSTSVVKGPGFTGG 496
QY 231 GVAIVPTKGQVSPVSGKIVVAFPSGHAFVRTKAEDGSNVVDILMHIGFTVNLNGTHFN 290
Db 497 DILRITSPGQISTL---RVITAPLSQRYVRIRY--ASTTNLQFHTSIDGRPINQGNFS 551
QY 291 PLKKQDEVKAGELLCEDFIDAIAKAGYEVVTPIVVSNYKKTGPVNYGLGEIAGANLL 350
Db 552 ATMSSGGNLQSG-----SFRTAGF--TTPF---NFSNGSSIETLSAHVFNNSG-NEV 596
QY 351 NVAKKEAVPA 360
Db 597 YIERIEFVPA 606

RESULT 7

5188960-2
Patent No. 5188960
APPLICANT: PAYNE, JEWEL; SICK, AUGUST J.
TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATE ACTIVE
AGAINST LEPIDOPTERAN PESTS, AND GENES ENCODING NOVEL
LEPIDOPTERAN-ACTIVE TOXINS
NUMBER OF SEQUENCES: 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/451,261
FILING DATE: 14-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 371,955
FILING DATE: 27-JUN-1989
SEQ ID NO: 2:
LENGTH: 1179
5188960-2

Query Match 5.8%; Score 106.5; DB 6; Length 1179;
Best Local Similarity 22.4%; Pred. No. 0.042;
Matches 56; Conservative 42; Mismatches 97; Indels 55; Gaps 13;

QY 143 MFLVCVVVTFIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG-----TTK 188
Db 380 LFVLDGTEFSFASLTADLPSTIY--RQGTVDLSLDVIPDQNSVVPARAGFSHRLSHVTML 437
QY 189 AEA-----EAP-----AEFSN---DSTIIQAPLTGEAIALSSVSDAMFASGKLGS 230

Db 438 SOAGAVYTLRAPTESWRHSAEFLNIPSSQITQIPLT-KSINLGSCTSVVKGPGFTGG 496
Qy 231 GVAIVPTKQGVSPVSGKIIVAFPSGHAFVTRKADSGSNVDILMHIGFDVNLNGTHFN 290
Db 497 DILRRTSPQISTL---RVITAPLSQRYRVRIRY--ASTNLQFHTSIDGRPINQGNFS 551
Qy 291 PLKKGQDEVKAGELLCEDIDAIKAAGYEVVTPPIVVSNNKKTGPNVTYGLGIEAGANLL 350
Db 552 ATMSSGGNLQSG-----SFRTAGF--TTPF---NFSNGSSIFTLSAHVNSG-NEV 596
Qy 351 NVAKKEAVPA 360
Db 597 YIDRIEFVPA 606

RESULT 8
US-08-894-818B-3
; Sequence 3, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894.818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.
US-08-894-818B-3

Query Match 5.5%; Score 99.5; DB 4; Length 522;
Best Local Similarity 20.6%; Pred. No. 0.063;
Matches 81; Conservative 63; Mismatches 114; Indels 135; Gaps 21;

Qy 1 YDFGPGVGLLGLVYSPVITGLHQSPPPIELEL-----FNQGSFIFAT-----ASMAN 51
Db 19 WNLGVDGSGITIG-----IIDTGIDASHPDLOGKVIWVDFVNGRSYPYDDHGHCHTHVAS 73
Qy 52 TAQGAACLAVEFFLAKSEKLKGLA-----GASGVSAVLGITEPAI-----FG 92
Db 74 TAAGTGA-----ASNGKYKGMAPCAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 127
Qy 93 ---VNLRLWPFFIGI-----GTAAGGALIALFNKAVALGAAGFLG-----132
Db 128 IKVINLSL-----GSSQSSDGTALSOAVNAADAGLVVVVAAGNSGPNKYITIGSPAAA 181
Qy 133 ---VVSIDAPDVMFLVCVAVTFFIAFG-----AAIAYGLYLV---RRNGS-----ID 174
Db 182 SKVITVGAVDKY-----DVITFSRSRGTADGRLKPEVVVAPGNWIIAARASGTSMGQPIN 236
Qy 175 PDATAAP-----VP---AGTTKRAEAEPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGK 227
Db 237 DYYTAAPGTSMATPHVAGIAALLLQAHPSWTPDKVKTALLETADIVKPDDEIADIAYAGR 296
Qy 228 LSGSVAI-----VPTKGQVSPVSGKIVVAFPPSGHAFVTRKADSGSNVDIL 274
Db 297 VNAYKAINVDNYAKLVFTGYVANKGS---QTHQFVI---SGASEFTATLYWDNANSOLD 349
Qy 275 MHI-----GEDTVNLNGTHNP 291
Db 350 LYLDPNGNQVDYSYATYGFKEVG-----YYPN 378

RESULT 9
US-08-894-818B-35
; Sequence 35, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894.818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-35

Query Match 5.5%; Score 99.5; DB 4; Length 654;
Best Local Similarity 20.8%; Pred. No. 0.089;
Matches 81; Conservative 63; Mismatches 114; Indels 135; Gaps 21;

QY 1 YDGGPVGGLFLGLVSPVITGLHOSFPPIEL-----FNOGGGFIPAT-----ASMAN 51
DB 151 WNLGYDGSGITIG-----IIDTGIDASHPLDGKVGWDFVNGRSYPYDDHGHGTHVAS 205
QY 52 IAOGAACLA VFFLAKSEKLGLA-----GASGVSAVLGITEPAI-----FG 92
DB 206 IAAGTGA-----ASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 259
QY 93 ---VNLRLRPFFIGI-----GTAATGGALIALFNKIKAVLAGAAGFLG----- 132
DB 260 IKVNLISL-----GSSQSDGTDALSOAVNAWDAGLVVVAAGNSGPNKYITIGSPAA 313
QY 133 --VVSIDAPDMVFLVCVVTFEIAFG-----AAIAYGLYLV--RNRGS-----ID 174
DB 314 SKVITGVNDKY-----DVITSSRGPTADGLKEVAVPAGWNIIAARASGTSMGQPIN 368
QY 175 PDATAAP-----VP--AGTTKABAEAPAEFSNDSTIIQAPLTGEAIALSVSDAMFASGK 227
DB 369 DYITAAPGTSMAHPHAGIAALLQAHPSTWPKVKTALETADIYKPDIEADIYAGAGR 428
QY 228 LGSQVAI-----VPTKGQLVSPVSKIVVAFPSGHAFVTRKAEDGSNVDIL 274
DB 429 VNAYKAINVDNYAKLVFTGVANKGS-----QTHQFVI-----SGASFVTATLYNDNANSLED 481
QY 275 MHI-----GFDVNLNGTHFNP 291
DB 482 LYLDPNQNOVDYSYAYYGFKEVG-----YYNP 510

RESULT 10
5254799-6
; Patent No. 5254799
; APPLICANT: DeGrove, Henri M. J.; Salgado, Maria B. L.; Van
; Montagu, Marc C. E.; Vaecck, Mark A.; Zabeau, Maarcus F. O.; Leemans,
; Jan J. A.; Hofte, Hermanus F. P.
; TITLE OF INVENTION: TRANSFORMATION VECTORS ALLOWING EXPRESSION
; OF BACILLUS THURINGIENSIS ENDOTOXINS IN PLANTS
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/555,828
; FILING DATE: 23-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 821,582
; FILING DATE: 22-JAN-1986
; APPLICATION NUMBER: 692,759
; FILING DATE: 18-JAN-1985
; SEQ ID NO: 6;
; LENGTH: 1184
5254799-6

Query Match 5.4%; Score 98; DB 6; Length 1184;
Best Local Similarity 19.9%; Pred. No. 0.33;
Matches 77; Conservative 44; Mismatches 127; Indels 138; Gaps 18;

QY 17 SPVITGLHQSPPIELELFNQGGSFIFATASMANIAQGAACLA VFFLAKSEKLGLAGA 76
DB 324 SPVGSFGEFAFP-----LFGNAGN-----AAPQORIAGNAA----- 355

QY 77 SGVSAVLGITEPAIFGVNLRWPFPIGICTAAIGGALIALFNKIKAVLAGAAGFLGVVSI 136
DB 356 ----PVLSTGLGIFRT---LSSPLY-----RRILGS----- 381
QY 137 DAPDMVFLVCVVTFEIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG----- 185
DB 382 -GPNQOEFSLVDGTEF--SFASLTTPNSTIYRGVTVDSLDVEIPQDNSVPPRAGFSHR 438
QY 186 -----TTKAEAEAPAEFSN---DSTIIQAPLTGEAIALSVSDAMFA 224
DB 439 LSHVTMLSOAGAVVYTLRATFSWQHSRAEFNNIIPSSQITQIPLT-----K 484
QY 225 SGKLGSGVAIIP-----TKGOLVSPVSG-----KIVVAFPSGHAFVTRKAEDGSNVDIL 274
DB 485 STNLGSGTSVVKPGPTGGHILARTSPGOISTLRVNITAPLSORYVRIRY--ASTNLQ 542
QY 275 MHIGFDTVNLNGTHFNPLKQGDVEKAGELLCEFDIDA KAAAGYEVTTPIVVSNYKKTGP 334
DB 543 FHTSIDGRPINOGNFSATMSSGSLNLSQSG-----SFTVGF--TTPP---NFSNGSS 588
QY 335 VNTYGLGEIEAGANLLNVAKKEAVPA 360
DB 589 VFTLSAHVNSG-NEVIIDRIEEVPA 613

RESULT 11
5254799-7
; Patent No. 5254799
; APPLICANT: DeGrove, Henri M. J.; Salgado, Maria B. L.; Van
; Montagu, Marc C. E.; Vaecck, Mark A.; Zabeau, Maarcus F. O.; Leemans,
; Jan J. A.; Hofte, Hermanus F. P.
; TITLE OF INVENTION: TRANSFORMATION VECTORS ALLOWING EXPRESSION
; OF BACILLUS THURINGIENSIS ENDOTOXINS IN PLANTS
; NUMBER OF SEQUENCES: 42
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/555,828
; FILING DATE: 23-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 821,582
; FILING DATE: 22-JAN-1986
; APPLICATION NUMBER: 692,759
; FILING DATE: 18-JAN-1985
; SEQ ID NO: 7;
; LENGTH: 1188
5254799-7

Query Match 5.4%; Score 98; DB 6; Length 1188;
Best Local Similarity 19.9%; Pred. No. 0.33;
Matches 77; Conservative 44; Mismatches 127; Indels 138; Gaps 18;

QY 17 SPVITGLHQSPPIELELFNQGGSFIFATASMANIAQGAACLA VFFLAKSEKLGLAGA 76
DB 324 SPVGSFGEFAFP-----LFGNAGN-----AAPQORIAGNAA----- 355
QY 77 SGVSAVLGITEPAIFGVNLRWPFPIGICTAAIGGALIALFNKIKAVLAGAAGFLGVVSI 136
DB 356 ----PVLSTGLGIFRT---LSSPLY-----RRILGS----- 381
QY 137 DAPDMVFLVCVVTFEIAFGAAIAYGLYLVRNGSID-----PDATAAPVPAG----- 185
DB 382 -GPNQOEFSLVDGTEF--SFASLTTPNSTIYRGVTVDSLDVEIPQDNSVPPRAGFSHR 438
QY 186 -----TTKAEAEAPAEFSN---DSTIIQAPLTGEAIALSVSDAMFA 224
DB 439 LSHVTMLSOAGAVVYTLRATFSWQHSRAEFNNIIPSSQITQIPLT-----K 484
QY 225 SGKLGSGVAIIP-----TKGOLVSPVSG-----KIVVAFPSGHAFVTRKAEDGSNVDIL 274
DB 485 STNLGSGTSVVKPGPTGGHILARTSPGOISTLRVNITAPLSORYVRIRY--ASTNLQ 542
QY 275 MHIGFDTVNLNGTHFNPLKQGDVEKAGELLCEFDIDA KAAAGYEVTTPIVVSNYKKTGP 334

COUNTRY: US
ZIP: 53701-2113
COMPUTER: IBM PC compatible
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/810,720
FILING DATE: 03-MAR-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 670513.90163
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608/251-5000
TELEFAX: 608/251-9166
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 648 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-810-720-13

Query Match 5.3%; Score 96.5; DB 3; Length 648;
Best Local Similarity 22.4%; Pred. No. 0.18;
Matches 53; Conservative 33; Mismatches 78; Indels 73; Gaps 13;
QY 166 LVRRNGSID-----PDATAAPVPAG-----TTKAEA-----EAP-----AE 196
Db 401 IYRQGTVDSDVIPPQDNSVPPRAGFSLHSHVMTLSQAAGAVYTLRAPTFQWHSRAE 460
QY 197 FSN---DSTTIQAPLTGEATLSVSDAMFASGKLGSGVAIVP-----TKQLVSPVSG-- 247
Db 461 FNNIIPSSQITQIPLT-----KSTNLGSGTSVVKGPFGTGGDILRTSPGQ 506
QY 248 ----KIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLGTHFNPLKKQGEVKAGE 303
Db 507 ISTRVNITAPLSQRYVRIRY--ASTNLQFHTSIDGRPINQGNFSATMSSGNSLNQSG- 563
QY 304 LLCEFDIDAIAKAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKEAVPA 360
Db 564 -----SPRTVGF--TTPF---NFSNGSSVFTLSAHVFNNSG-NEVIIDRIEVPVA 606

RESULT 15
US-08-446-486-7
Sequence 7, Application US/08446486
Patent No. 5545565
GENERAL INFORMATION:
APPLICANT: De Greve, Henri Marcel J
APPLICANT: Salgado, Maria Benita Leonor F
APPLICANT: Van Montagu, Marc Charles E
APPLICANT: Vaeck, Mark A
APPLICANT: Zabeau, Marcus Florent O
APPLICANT: Leemans, Jan Jozef A
APPLICANT: Hofte, Hermanus Franciscus P
TITLE OF INVENTION: TRANSMFORMATION VECTORS ALLOWING
TITLE OF INVENTION: EXPRESSION OF FOREIGN POLYPEPTIDE ENDOTOXINS FROM BACILLUS
TITLE OF INVENTION: THURINGIENSIS IN PLANTS
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: George Mason Bldg., Washington & Prince Sts.
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,486
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: US 08/133,965
FILING DATE: 08-OCT-1993
APPLICATION NUMBER: US 08/014,148
FILING DATE: 05-FEB-1993
PRIOR APPLICATION DATA: US 07/555,828
FILING DATE: 23-JUL-1990
PRIOR APPLICATION DATA: US 06/821,582
FILING DATE: 22-JAN-1986
PRIOR APPLICATION DATA: US 06/692,759
FILING DATE: 18-JAN-1985
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-047
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..934
OTHER INFORMATION: /note= "deduced amino acid sequence
OTHER INFORMATION: of B.t. sotto (Shibano et al., Gene 34, p. 243,
OTHER INFORMATION: 1985)"
US-08-446-486-7

Query Match 5.3%; Score 96.5; DB 1; Length 934;
Best Local Similarity 22.4%; Pred. No. 0.32;
Matches 53; Conservative 33; Mismatches 78; Indels 73; Gaps 13;
QY 166 LVRRNGSID-----PDATAAPVPAG-----TTKAEA-----EAP-----AE 196
Db 401 IYRQGTVDSDVIPPQDNSVPPRAGFSLHSHVMTLSQAAGAVYTLRAPTFQWHSRAE 460
QY 197 FSN---DSTTIQAPLTGEATLSVSDAMFASGKLGSGVAIVP-----TKQLVSPVSG-- 247
Db 461 FNNIIPSSQITQIPLT-----KSTNLGSGTSVVKGPFGTGGDILRTSPGQ 506
QY 248 ----KIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLGTHFNPLKKQGEVKAGE 303
Db 507 ISTRVNITAPLSQRYVRIRY--ASTNLQFHTSIDGRPINQGNFSATMSSGNSLNQSG- 563
QY 304 LLCEFDIDAIAKAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKEAVPA 360
Db 564 -----SPRTVGF--TTPF---NFSNGSSVFTLSAHVFNNSG-NEVIIDRIEVPVA 606

Search completed: March 21, 2002, 16:23:18
Job time: 149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 21, 2002, 16:22:45 ; Search time 42.58 Seconds

(without alignments)
629.745 Million cell updates/sec

Title: US-09-604-231-4

Perfect score: 1821

Sequence: 1 YDFGPGVGLLFLGVSPV.....IEAGANLLNVAKKAVPATP 362

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

```
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1821	100.0	362	22	AAB66708 C. glutamicum phosph
2	1821	100.0	468	22	AAB66707 C. glutamicum phosph
3	1821	100.0	661	22	AAG92650 C. glutamicum prote
4	1806	99.2	661	22	AAB69080 Brevibacterium lac
5	390.5	21.4	683	22	AAG93207 C. glutamicum prote
6	390.5	21.4	683	22	AAB66721 C. glutamicum phosph
7	295	16.2	648	22	AAG98282 Escherichia coli p
8	234	12.9	135	22	AAB66723 C. glutamicum phosph
9	233.5	12.8	167	20	AA120012 B. burgdorferi ant
10	221.5	12.2	583	21	AA153620 Amino acid sequenc
11	166.5	9.1	135	20	AA120013 B. burgdorferi ant

12	161.5	8.9	126	18	AAW28051	Amino acid sequenc
13	154	8.5	108	21	AAB40791	Human OREF ORF555
14	150.5	8.3	449	21	AA181556	Streptococcus pneu
15	118.5	6.5	805	22	AAG90449	C. glutamicum prote
16	113.5	6.2	1176	14	AA139755	Delta endotoxin.
17	112	6.2	1951	22	AA199678	Human adult form o
18	112	6.2	1951	22	AA199679	Human neonatal for
19	108.5	6.0	806	20	AA121797	Alteromonas lipase
20	108.5	6.0	809	20	AA121796	Alteromonas lipase
21	106.5	5.8	1179	11	AA18258	B. thuringiensis to
22	106.5	5.8	1179	13	AA125826	Novel toxin expres
23	105	5.8	404	22	AA191953	C. glutamicum prote
24	105	5.8	1291	19	AA159912	Amino acid sequenc
25	102.5	5.6	474	19	AA158862	T. halophilus xyl
26	102	5.6	655	22	AA198089	C. glutamicum prote
27	101.5	5.6	405	22	AA15136	Peptide #1570 enco
28	101.5	5.6	405	22	AA127591	Peptide #1628 enco
29	101.5	5.6	405	22	AA12877	Peptide #1559 enco
30	101.5	5.6	405	22	AA14075	Peptide #2757 enco
31	101.5	5.6	2009	22	AA199674	Human adult form o
32	101	5.5	521	21	AA13578	Streptomyces globi
33	101	5.5	521	21	AA13605	Streptomyces globi
34	100.5	5.5	459	12	AA13498	P. denitrificans CO
35	100.5	5.5	697	22	AA178983	C. glutamicum SRT
36	100.5	5.5	772	22	AA192907	C. glutamicum prote
37	99.5	5.5	412	20	AA194836	Hyperthermostable
38	98.5	5.5	522	18	AA124122	Pyrococcus furiosu
39	99.5	5.5	522	20	AA194838	Hyperthermostable
40	99.5	5.5	654	18	AA124129	Pyrococcus furiosu
41	99.5	5.5	654	20	AA194841	Hyperthermostable
42	99	5.4	1272	21	AA170596	Rat multidrug resi
43	99	5.4	1272	21	AA170597	Rat multidrug resi
44	98.5	5.4	440	20	AA100194	Enterococcus faeca
45	98.5	5.4	2005	22	AA199676	Human adult form o

ALIGNMENTS

RESULT 1
AAB66708
ID AAB66708 standard; protein; 362 AA.
XX
AC AAB66708;
XX
DT 09-APR-2001 (first entry)
XX
DE C. glutamicum phosphoenolpyruvate protein #2.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB009973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
(BADI) BASF AG.
Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;
WPI; 2001-080989/09.
Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;
sugar phosphotransferase system proteins or their portions, useful for
typing or identifying C. glutamicum or related bacteria, and as markers

```

PT for transformation -
XX
PS Claim 6; Page 104-106; 144pp; English.
XX
XX The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation..
XX
SQ Sequence 362 AA;

Query Match 100.0%; Score 1821; DB 22; Length 362;
Best Local Similarity 100.0%; Pred. No. 6.5e-171;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YDFGGPVGGLLGLVSPVITGLHOSFPPIELFENOGGSFIFATASMANIAQGAACLA 60
DB 1 Ydfggpvvgllglvsvpivitglhgsfppllelfnqggsfifatasmaniaqgaacia 60
QY 61 VFELAKSEKLGLAGASGVSALVIGITEPAIFGVNLRWPFFIGIGTAAIGGALIALFNI 120
DB 61 vfelakseklglagagsvsavligitepaifgvnlrlrwpffigigtaaiaggallalfni 120
QY 121 KAVALGAAGFLGVVSDADPMVMFLVCVVTFPIAFGAATAYGLYLVRNNGSIDPDATAA 180
DB 121 kavalgaagflgvvsidapdmvmflvcavvtfpiafgaataayglylvrrnngsidpdataa 180
QY 181 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
DB 181 pvpagttkaeaeapaeFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
QY 241 LVSPVSGKIVWAPPSGHAFVTRTKAEDGNSVDILMHIGFTVNLNGTHFNPLKKQGDVKK 300
DB 241 lvspvsgkivwappsgghafvtrtkaedgnsvdilmhigftvnlngthfnplkkqgdevk 300
QY 301 AGEELCEFDIDAIAKAAGYEVVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 360
DB 301 ageelcefdidaikaagyeVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 360
QY 361 TP 362
DB 361 tp 362

RESULT 2
AAB66707
ID AAB66707 standard; protein; 468 AA.
XX
AC AAB66707;
XX
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate protein #1.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
XX
PR 23-AUG-1999; 99US-0150310.
XX
PR 03-SEP-1999; 99DE-1042095.
XX
PR 03-SEP-1999; 99DE-1042097.

```

```

XX (BADI ) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI: 2001-080989/09.
XX
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
XX sugar phosphotransferase system proteins or their portions, useful for
XX typing or identifying C. glutamicum or related bacteria, and as markers
XX for transformation -
XX
XX Claim 4; Page 101-102; 144pp; English.
XX
XX The present invention relates to Corynebacterium glutamicum
XX phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
XX The PTS nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria, the typing or identification of C. glutamicum or
XX related bacteria, as reference points for mapping C. glutamicum genome,
XX and as markers for transformation.
XX
SQ Sequence 468 AA;

Query Match 100.0%; Score 1821; DB 22; Length 468;
Best Local Similarity 100.0%; Pred. No. 9.4e-171;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YDFGGPVGGLLGLVSPVITGLHOSFPPIELFENOGGSFIFATASMANIAQGAACLA 60
DB 107 ydfggpvvgllglvsvpivitglhgsfppllelfnqggsfifatasmaniaqgaacia 166
QY 61 VFELAKSEKLGLAGASGVSALVIGITEPAIFGVNLRWPFFIGIGTAAIGGALIALFNI 120
DB 167 vfelakseklglagagsvsavligitepaifgvnlrlrwpffigigtaaiaggallalfni 226
QY 121 KAVALGAAGFLGVVSDADPMVMFLVCVVTFPIAFGAATAYGLYLVRNNGSIDPDATAA 180
DB 227 kavalgaagflgvvsidapdmvmflvcavvtfpiafgaataayglylvrrnngsidpdataa 286
QY 181 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
DB 287 pvpagttkaeaeapaeFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 346
QY 241 LVSPVSGKIVWAPPSGHAFVTRTKAEDGNSVDILMHIGFTVNLNGTHFNPLKKQGDVKK 300
DB 347 lvspvsgkivwappsgghafvtrtkaedgnsvdilmhigftvnlngthfnplkkqgdevk 406
QY 301 AGEELCEFDIDAIAKAAGYEVVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 360
DB 407 ageelcefdidaikaagyeVTPPIVVSNNYKKTGPVNTYGLGTEAGANLLNVAKKEAVPA 466
QY 361 TP 362
DB 467 tp 468

RESULT 3
AAG92650
ID AAG92650 standard; Protein; 661 AA.
XX
AC AAG92650;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6404.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.

```

```
XX EP1108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR
XX
XX 07-APR-2000; 2000JP-0159162.
PR
XX
XX 03-AUG-2000; 2000JP-0280988.
PR
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR
XX
XX N-PSDB; AAH67869.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 661 AA;
SQ
Query Match 100.0%; Score 1821; DB 22; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.6e-170;
Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YDFGPGVGLLFGVYSPVITGLHQSPPIELELFNNGGSPFIFATASMANIAQGAACIA 60
DB 300 ydfgpgvgllfglvyspivltglhqsfpplielelfnngggsfifatasmanlaqgaacia 359
QY 61 VFFLAKSEKLKGLAGASGVSAVLGTETPAIFGVNRLRWPFPIGIGTAIGGALLIALFNI 120
DB 360 vfflakseklkglagsgvsavlgitepaifgvnlrlrwpffigigtaiaggallialfni 419
QY 121 KAVALGAAGFLGVVSDPADPMVMFLVCVAVVFFIAFGAAIAYGLVLRNCSIDPDATAA 180
DB 420 kavalgaagfglgvvsidpadpmvmflvcavvttfifaagaiaayglvlvrnsgidpdataa 479
QY 181 PVPAGTTKAEAEAPAEFNSDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
DB 480 pvpagttkaeaeapaeafnsdstiiqapltgeaialssvsdamfasklgsgvaivptkgq 539
QY 241 LVSPVSGKIVVAFPSGSHAFVTRKAEKDSNVDILMHIGFDTVNLNTHFNPLKKGDEVK 300
DB 540 lvspvsgkivvafpsghafvtrkaekdsnvdiilmhigfdtnlnthfnplkkgdevk 599
QY 301 AGELLCEFDIDAIRKAGYEVTPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPA 360
DB 600 agellcefdidaikagyevttpivvsnnkktgpvntyglgeieaganllnvakeavpa 659
QY 361 TP 362
XX
||
```

```
Db 660 tp 661
RESULT 4
AAB69080
ID AAB69080 standard; Protein; 661 AA.
XX
XX AAB69080;
AC
XX 20-APR-2001 (first entry)
DT
XX
XX Brevibacterium lactofermentum sucrose PTS enzyme II protein SEQ ID NO:2.
DE
XX
XX Brevibacterium lactofermentum: sucrose PTS enzyme II; sucrose;
KW phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;
KW coryneform bacterium; phosphoenolpyruvate-sugar transport system.
XX
XX Brevibacterium lactofermentum.
OS
XX
XX WO200102584-A1.
PN
XX 11-JAN-2001.
PD
XX
XX 30-JUN-2000; 2000WO-JP04348.
PF
XX
XX 02-JUL-1999; 99JP-0189512.
PR
XX
XX (AJIN ) AJINOMOTO CO INC.
PA
XX
XX Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;
PI
XX
XX WPI; 2001-138150/14.
DR
XX
XX N-PSDB; AAF32543.
XX
XX Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme
PT II obtained by cassette ligation-mediated amplification of downstream
PT domain of coryneform bacterium sucrose gene, with sucrose-binding
PT activity -
XX
XX Claim 1; Page 29-32; 45pp; Japanese.
PS
XX
XX The present sequence represents the Brevibacterium lactofermentum
CC sucrose PTS (phosphoenolpyruvate:carboxylate phosphotransferase system
CC or phosphoenolpyruvate-sugar transport system) enzyme II, which has
CC sucrose-binding activity. A coryneform bacteria produced with the
CC sucrose PTS enzyme II gene can have more efficient sugar uptake, and
CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene
CC and it's disrupted gene, such as one without the sucrose PTS function,
CC can be used to produce new breeds of coryneform bacterial strains to
CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can
CC have improved amino-acid and nucleic acid productivity.
XX
XX Sequence 661 AA;
SQ
Query Match 99.2%; Score 1806; DB 22; Length 661;
Best Local Similarity 99.2%; Pred. No. 4.7e-169;
Matches 359; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 YDFGPGVGLLFGVYSPVITGLHQSPPIELELFNNGGSPFIFATASMANIAQGAACIA 60
DB 300 ydfgpgvgllfglvyspivltglhqsfpplielelfnngggsfifatasmanlaqgaacia 359
QY 61 VFFLAKSEKLKGLAGASGVSAVLGTETPAIFGVNRLRWPFPIGIGTAIGGALLIALFNI 120
DB 360 vfflakseklkglagsgvsavlgitepaifgvnlrlrwpfigigtaiaggallialfai 419
QY 121 KAVALGAAGFLGVVSDPADPMVMFLVCVAVVFFIAFGAAIAYGLVLRNCSIDPDATAA 180
DB 420 kavalgaagfglgvvsidpadpmvmflvcavvttfifaagaiaayglvlvrnsgidpdataa 479
QY 181 PVPAGTTKAEAEAPAEFNSDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 240
XX
||
```


Db 480 pvpagttkaaeapafesndstiiqapltgeaialssvsdamfasgklgsqvaivptkqg 539
QY 241 LVSPVSGKIIVAPSGHAFVRKAEDGSNVDLIMHIGFDTVNLNGTHFNPLKKQGDVYK 300
Db 540 lvspvsgkiivafpsghafavrkaedgsnvdlimhigfdtnlngthfnplkkqgdevk 599
QY 301 AGELLCEFDIDAIAKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 360
Db 600 agellcefdidaikaagyevttipivsnkyktpvntyglgeieaganllnvaakkeavpa 659
QY 361 TP 362
Db 660 tp 661

RESULT 5
AAG93207
ID AAG93207 standard; Protein; 683 AA.
XX
AC AAG93207;
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6961.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000BP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
DR N-PSDB; AAH68426.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
PS Claim 29; SEQ ID NO: 6961; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 683 AA;

Query Match

21.4%; Score 390.5; DB 22; Length 683;

Best Local Similarity 30.7%; Pred. No. 1.3e-29;
Matches 118; Conservative 60; Mismatches 159; Indels 47; Gaps 9;
QY 10 LLFGLVYSPVITVGLHQSPPIELFNLQ-GGSFIFATASMANIAQGAACLAFFFLAKSE 68
Db 307 iviplypflvplghwplnaimlqntlgydfggpmgawnfacfglvtgfvllsike 366
QY 69 KLKGLAGAS--GVSAYL--GITEPATFGVNLRLRWPFIFIGTAAIGGALIALFNKAVA 124
Db 367 rnkamrqvslggmlagllggiseplsygvlrfrkkyfrllp9cslagglvimgifdkaya 426
QY 125 LGAAGFLGVVSDIDAPDMVMFLVCAVVTFFIAFGAAIAYGLYLVRNRGID----- 174
Db 427 fvftslitipamd--pwlgytigiavaffvsmflviald--yrsneerdearakvaadk 481
QY 175 -----PDAT-AAPV-----PAGTTKAEAEAPESNDSTIIQAPLTGE 211
Db 482 qaeedikaeanatpaapvaaagagagagagaagaaatavaakpklaagevvdvpslegk 541
QY 212 AIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIVVAFPSGHAFVRTRKEDGSNV 271
Db 542 aiplsevpdpifaagklgpgiaiqptgntvvpadatcvilivqksghavairl-----ds9v 597
QY 272 DILMHIGFDTVNLNGTHFNPLKKQGDVYKAGELICEFDIDAIAKAAGYEVTTPIVVSNYK 331
Db 598 eilvhvgldtvglggegftvhrqqvkgadplttfdadfirskdplltvpsvsnak 657
QY 332 TGPVNTYGLGEIEAGANLLNVAKK 355
Db 658 fgeiegipadqanssttvikvngk 681

RESULT 6
AAB66721
ID AAB66721 standard; protein; 683 AA.
XX
AC AAB66721;
DT 09-APR-2001 (first entry)
XX
DE C.glutamicum phosphoenolpyruvate protein #5.
XX
KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX
OS Corynebacterium glutamicum.
XX
PN WO200102583-A2.
XX
PD 11-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-IB00973.
XX
PR 01-JUL-1999; 99US-0142691.
PR 23-AUG-1999; 99US-0150310.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042097.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI: 2001-080989/09.
XX
PT Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation
XX
PS Claim 32; Page 136; 144pp; English.
XX
CC The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the


```

XX AAB66723;
AC
XX
XX 09-APR-2001 (first entry)
DT
XX
XX C.glutamicum phosphoenolpyruvate protein #7.
DE
XX
XX Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
KW
XX
XX Corynebacterium glutamicum.
XX
XX W0200102583-A2.
XX
XX 11-JAN-2001.
XX
XX 27-JUN-2000; 2000WO-1B00973.
XX
XX 01-JUL-1999; 99US-0142691.
XX
XX 23-AUG-1999; 99US-0150310.
XX
XX 03-SEP-1999; 99DE-1042095.
XX
XX 03-SEP-1999; 99DE-1042097.
XX
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-080989/09.
XX
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
XX
XX sugar phosphotransferase system proteins or their portions, useful for
XX
XX typing or identifying C. glutamicum or related bacteria, and as markers
XX
XX for transformation -
XX
XX Claim 36; Page 142-143; 144pp; English.
XX
XX The present invention relates to Corynebacterium glutamicum
XX
XX phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.
XX
XX The PTS nucleic acids and proteins are useful in the
XX
XX identification of microorganisms which can be used to produce fine
XX
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX
XX related bacteria, the typing or identification of C. glutamicum or
XX
XX related bacteria, as reference points for mapping C. glutamicum genome,
XX
XX and as markers for transformation.
XX
XX Sequence 135 AA;

Query Match 12.9%; Score 234; DB 22; Length 135;
Best Local Similarity 37.5%; Pred. No. 3.1e-15;
Matches 51; Conservative 25; Mismatches 56; Indels 4; Gaps 1;

QY 220 DAMFASGKLGSGVAIVPTKGOLVSPVSGKIVAFPSGHAFVTRKAEDGSNVDILMHIGF 279
Db 2 dpifagkgipggaigqptgntvvpadatvllvqkghavalfi----dsgveillvhvg 57

QY 280 DTYNLNGTHFNPLKKGQDEVKAGELLCFEDDAIKAAAGVEVTPPIVVSNNYKTKGPVNTYG 339
Db 58 dtvqlgqegftvhvrrqvkagdpiltdfadirskdplitpvtvvsnaakfgeiegip 117

QY 340 LGEIEAGANLLNVAKK 355
Db 118 adqanssttvkvngk 133

RESULT 9
AAY20012
ID AAY20012 standard; Protein; 167 AA.
XX
XX AAY20012;
XX
XX 19-JUL-1999 (first entry)
DT
XX
XX B. burgdorferi antigenic protein, f346.aa.

```

```

XX Antigenic protein; vaccine; Lyme disease; infection; detection.
KW
XX
XX Borrelia burgdorferi.
OS
XX
XX W09859071-A1.
XX
XX 30-DEC-1998.
XX
XX 18-JUN-1998; 98WO-US12718.
XX
XX 03-SEP-1997; 97US-0057483.
XX
XX 20-JUN-1997; 97US-0050359.
XX
XX 22-JUL-1997; 97US-0053344.
XX
XX 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX (MEDI-) MEDIMMUNE INC.
XX
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
XX WPI; 1999-189980/16.
XX
XX N-PSDB; AAX61709.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX
XX products for the diagnosis, prevention and treatment of diseases
XX
XX caused by Borrelia, particularly Lyme disease
XX
XX Claim 12; Page 161; 275pp; English.
XX
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX
XX can be used in vaccines for eliciting protective antibodies to members of
XX
XX the Borrelia genus, particularly for the use against Lyme disease in
XX
XX humans and animals. They can be used for preventing or attenuating an
XX
XX infection caused by a member of the Borrelia genus. The products can also
XX
XX be used for detection of members of the Borrelia genus.
XX
XX Sequence 167 AA;

Query Match 12.8%; Score 233.5; DB 20; Length 167;
Best Local Similarity 34.8%; Pred. No. 4.7e-15;
Matches 46; Conservative 30; Mismatches 51; Indels 5; Gaps 2;

QY 213 IALSSVSDAMFASGKLGSGVAIVPTKGOLVSPVSGKIVAFPSGHAFVTRKAEDGSNVD 272
Db 1 msidkvpdeafaekivgdgiailptsnellapcdgkigkiftnhafsletk----egve 56

QY 273 ILMHIGFDTYNLNGTHFNPLKKGQDEVKAGELLCFEDDAIKAAAGVEVTPPIVVSNNYKKT 332
Db 57 ifvhfgintlnlgkftvraeeeginvkqgeviiridleylkehsevitpvtviansdev 116

QY 333 GPVNTYGLGEIE 344
Db 117 ssie-ysfgrle 127

RESULT 10
AAY53620
ID AAY53620 standard; Protein; 583 AA.
XX
XX AAY53620;
XX
XX 22-FEB-2000 (first entry)
DT
XX
XX Amino acid sequence of a permease encoded by the lactose operon.
XX
XX Lactose operon; permease; beta-galactosidase; mutant; yoghurt;
XX
XX beta-galactosidase activity; lactose fermentation;
XX
XX fermented milk product; Streptococcus thermophilus.
XX
XX Lactobacillus bulgaricus.
OS

```



```

XX AC AAW28051;
XX DT 27-AUG-1998 (first entry)
XX DE Amino acid sequence of Staphylococcus aureus protein.
XX KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
XX KW Staphylococcal gene; regulatory element; bacterial gene expression;
XX KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
XX KW toxic shock syndrome.
XX OS Staphylococcus aureus.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1..126 /note= "residues designated X are not defined in
XX FT the specification"
XX PN W09730070-A1.
XX XX
XX PD 21-AUG-1997.
XX XX
XX PF 19-FEB-1997; 97WO-US02318.
XX XX
XX PR 20-FEB-1996; 96US-0011888.
XX XX
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX XX
XX PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
XX PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
XX XX
XX DR WPI: 1997-424969/39.
XX DR N-PSDB; AAT84002.
XX XX
XX PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
XX PT to isolate antimicrobial compounds, and in vaccines against S.
XX PT aureus infection
XX XX
XX PS Claim 6; Pages 432-433; 989pp; English.
XX XX
XX CC The present sequence represents a Staphylococcus aureus protein, that,
XX CC based on homology with a human or Staphylococcus aureus protein, is
XX CC believed to be a nerve growth factor or p15G protein. The DNA sequence
XX CC was isolated from a library of clones of S. aureus WCUH 29 in
XX CC Escherichia coli. The DNA sequence can be used in the construction of
XX CC ribozymes and antisense sequences to control the expression of
XX CC Staphylococcal genes. The DNA sequence is also useful as a source of
XX CC regulatory elements for the control of bacterial gene expression. The
XX CC present protein may be used to produce vaccines to enable a host to
XX CC produce specific antibodies with antibacterial action. These vaccines
XX CC and antibodies would protect a host against invasion by S. aureus, and
XX CC conditions relating to Staphylococcal infection, e.g. Staphylococcal
XX CC food poisoning, scaled skin syndrome, and toxic shock syndrome.
XX SQ Sequence 126 AA;

Query Match 8.9%; Score 161.5; DB 18; Length 126;
Best Local Similarity 32.8%; Pred. No. 3.8e-08;
Matches 40; Conservative 25; Mismatches 44; Indels 13; Gaps 4;

OY 228 LGSVVAI-VPTKQLVSPVSGKIVAVPSPGHAFVTKAEDGNSVDILMHIGFDTYNLNG 286
Db 2 vgeglaikheeskvlapngflismlyptkdhvgi--qsedg--vdvihgkvxvdleg 57
OY 287 THENPLKKGQGVKAGELCELFIDDAIKAAGYEVTTPIVVSNNYKKTGPVNTYGLGEIAG 346
Db 58 kgfkfyvngndhveaqtqlqfdqgyiqxgynadxivvis-----nsadlqkvelt 109
OY 347 AN 348
Db 110 mn 111

```

```

RESULT 13
AAB40791
ID AAB40791 standard; Protein; 108 AA.
XX
AC AAB40791;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF555 polypeptide sequence SEQ ID NO:1110.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX KW hypotensive; dermatological; immunosuppressive; antinflammatory;
XX KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX KW cholesterol ester storage; systemic lupus erythematosus; infection;
XX KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX KW thrombosis; contraceptive.
XX OS Homo sapiens.
XX XX
XX PN W0200058473-A2.
XX XX
XX PD 05-OCT-2000.
XX XX
XX PF 31-MAR-2000; 2000WO-US08621.
XX XX
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX XX
XX PA (CURA-) CURAGEN CORP.
XX XX
XX PI Shinkets RA, Leach M;
XX XX
XX DR WPI: 2000-602362/57.
XX DR N-PSDB; AAC75000.
XX XX
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX XX
XX PS Claim 11; Page 1038; 5507pp; English.
XX XX
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antirheumatic; coagulant; vasotropic;
XX CC immunostimulant; cardiant; thrombolytic; immunosuppressant;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antithyroid; and antianaemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antinflammatory disease; to enhance

```

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 108 AA;

Query Match

Best Local Similarity 8.5%; Score 154; DB 21; Length 108;

Matches 32; Conservative 19; Mismatches 43; Indels 4; Gaps 1;

QY 194 PAEFSNDSTIIQAPTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKIIVAF 253

Db 14 psslkpsskivaplggllvlpdqdpvfqckmvgdgisldpnsnellapvagtvtqlh 73

QY 254 PSGHAFVATKRAEDGSNDVILMHIGFDTVNLNGTHNP 291

Db 74 nanhaltitp-----egievlnhigdtvmlrgdsypp 107

RESULT 14

AAAY81556

ID AAY81556 standard; Protein; 449 AA.

XX AC AAY81556;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae type 4 protein sequence #56.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;

KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;

KW pneumococcal disease.

XX Streptococcus pneumoniae.

OS WO200006737-A2.

PN 10-FEB-2000.

PD 27-JUL-1999; 99WO-GB02451.

PF 27-JUL-1999; 98GB-0016337.

PR 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHNIKS LTD.

PA Gilbert CFG, Hansbro PM;

XX WPI; 2000-195300/17.

DR New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein

XX Claim 1; Page 78; 108pp; English.

PS AAY81501 to AAY81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AAY81501 to AAY81590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AAY81591 to AAY81614 represent primers used in the
CC exemplification of the present invention.

XX Sequence 449 AA;

Query Match

8.3%; Score 150.5; DB 21; Length 449;

Best Local Similarity 26.7%; Pred No. 3e-06;

Matches 56; Conservative 29; Mismatches 68; Indels 57; Gaps 7;

QY 7 VGGLLFGLVSPVITIGLHQSPFPIEIELEFNOGSGFIFATAS---MANIAQGAA----- 57

Db 255 iglmliylltlqilwivgh-----ganivfafvspialanmaenaagghfav 301

QY 58 -----CLAVFFLAKSEKLKGLAGASVAVLGITEPAIFGV----- 93

Db 302 agefsnmfviagsgatlgcllyafaskseqikaigrasvvpalfnneplifgipily 361

QY 94 NLRLRPFPIG-IGTAAIGGALIALFNKAV-----ALGAAGFLGVVSDAPDMVMF 144

Db 362 npalaipfilapmvvtaiyyvanslnfikpiaqvptpvgigafgtadira---vlva 419

QY 145 LVCVVVTFEAFGAATAYGLYLVRNGSID 174

Db 420 lvcataaflvlypfirvydqkivkeegiz 449

RESULT 15

AAG90449

ID AAG90449 standard; Protein; 805 AA.

XX AC AAG90449;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 4203.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS EP1108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

DR N-PSDB; AAH65668.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 4203; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

```
XX      Query Match
SQ      Sequence      805 AA;
      6.5%; Score 118.5; DB 22; Length 805;
      Best Local Similarity 22.8%; Pred. No. 0.0099;
      Matches 101; Conservative 55; Mismatches 174; Indels 113; Gaps 20;

QY      7 VGGLLFGLVYSPIVI-----TGLHQSF-----PPIELELEFNQGGSFIFATASMANIAQG- 55
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      174 lggapf-vvaspdalmewnsgrvgefrvatsdpasle-----aasfsdatvvvas-aegh 227

QY      56 -----AACIAVFFFLAKSEKLGKLAGAS-----GVSAVLGITEP 88
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      228 vdkladsylggrdcyfyllaafvavaaavflvvfsaysvltgervrefglirsvgastp 287

QY      89 AIFGVNLRRLWPPFIGI-GTAAIG-GALIALFNKIKAVALGAAGF---LGVVSDIDAPDMVM 143
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      288 qilgsvi-----feagilgvvaagfgapaglmaarlldnaarfgrirvpdivdlpsstm 342

QY      144 FLVCVVTFETAFGAATAYGLYVRRNGSIDPDATAA-----PVPAGTTKAEA 191
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      343 wliagvqvmsviaalpa--vfvscrksavesistpairstspwfgalwlllagivgag 400

QY      192 EAPAEFNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKQQLVSPVSGKIVV 251
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      401 mwayeatsdyrgmr-----svalsiaagsalvcalliatavlvpwlhvfrivggtvp 454

QY      252 AFPSGHAFVTRKAEKDSNVVDILM-----HIGFDTV-----NLNGTHFNPLK 293
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      455 tlgqlafafaakqrsaalilavilagsalsavlhqahighthlvavakmgggtddmmta 514

QY      294 KQGEVKAGELLCFEFDIDATKAAGYEVTTPIVVSNNYKKTGPVNTYGLGEIEAGANLL--- 350
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      515 ldg-eipagmleeissldgvktaipattaveled---sgnfsvlmlae-edgasvmrag 569

QY      351 -----NVAKKEAVPA 360
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      570 dtgapagglvlgrrnsdpqdaypa 592
```

Search completed: March 21, 2002, 16:22:47
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2002, 08:11:36 ; Search time 1607.29 Seconds

(without alignments)

7414.386 Million cell updates/sec

Title: us-09-604-231-3

Perfect score: 1109

Sequence: 1 tatgatctcggtccagt.....gttgaacctgaagtcttcg 1109

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	45.8	4.1	473	11	T03017 FB20C8 Feta
2	41.8	3.8	681	13	CNS02EOD
3	41.6	3.8	895	13	CNS0071A
4	41.4	3.7	477	11	BE997958 Drosophil
5	41.4	3.7	559	10	AW574083
6	41.4	3.7	601	11	BE997958
7	41.4	3.7	627	11	BE998413
8	40.8	3.7	884	13	CNS00600
9	40.8	3.7	983	13	CNS03CVD
C 10	40.2	3.6	970	13	CNS03H6V
11	39.8	3.6	606	11	BE997957
C 12	39.2	3.5	993	13	CNS00LOP

C 13	38.8	3.5	539	10	AI070146
14	38.8	3.5	714	10	AW583970
C 15	38.6	3.5	401	10	AI436816
16	38.6	3.5	1204	11	EG391833
17	38.4	3.5	1101	13	CNS0021D
C 18	38.2	3.4	525	13	CNS025EN
19	38.2	3.4	531	10	AW573719
20	38.2	3.4	621	11	BE998387
C 21	38.2	3.4	652	10	AI982977
22	38.2	3.4	687	10	AW980716
23	38.2	3.4	703	10	AW980735
24	38.2	3.4	713	11	EG581619
25	38.2	3.4	737	11	EG581892
26	38.2	3.4	789	11	EG582423
27	38.2	3.4	795	11	EG583630
C 28	37.8	3.4	654	10	AI812147
29	37.6	3.4	1101	13	CNS00FXE
C 30	37.4	3.4	460	10	AI318021
31	37.4	3.4	552	10	BE71200
C 32	37.4	3.4	904	13	CNS03HAO
33	37.2	3.4	421	10	AW173566
34	37.2	3.4	561	10	AI917973
35	37.2	3.4	594	10	AW440291
36	37.2	3.4	622	10	AI499080
37	37.2	3.4	650	10	AW516253
38	37.2	3.4	710	10	AI564728
39	37.2	3.4	906	13	CNS02GJN
C 40	37	3.3	1101	13	CNS01602
C 41	36.8	3.3	458	10	AU060899
C 42	36.8	3.3	642	11	BF599151
C 43	36.8	3.3	1159	13	CNS015XR
44	36.6	3.3	902	13	CNS006QP
C 45	36.6	3.3	1101	13	CNS017SY

ALIGNMENTS

RESULT 1

T03017/c

LOCUS FB20C8 Fetal brain, Stratagene Homo sapiens cDNA clone FB20C8, mRNA
DEFINITION T03017 473 bp
ACCESSION T03017
VERSION T03017.1
KEYWORDS GI:314258
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 473)
AUTHORS Khan, A.S., Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A., Stevens, T.J., Robinson, M., Orpana, A.K. and Sikela, J.M.
TITLE Single pass sequencing and physical and genetic mapping of human CDNA
JOURNAL Nature Genet. 2, 180-185 (1992)
MEDLINE 94258200
COMMENT On Sep 21, 1992 this sequence version replaced gi:279156.
Contact: Sikela JM
Department of Pharmacology
University of Colorado Health Sciences Center
Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
Tel: 3032708637
Fax: 3032707097
Email: nikkit@tally.uchsc.edu.
Location/Qualifiers
1. 473
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):86662"
/db_xref="GDB:D052785E"
/db_xref="taxon:9606"
/clone="FB20C8"
/clone_lib="Fetal brain, Stratagene"

14-FEB-1997

JOURNAL REFERENCE Unpublished
AUTHORS 2 (bases 1 to 983)
 Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 983)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.
FEATURES location/Qualifiers
 source
 1..983
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG015CB04SP1-end : PUC-ori"
BASE COUNT 202 a 289 c 274 g 201 t 17 others
ORIGIN
 Query Match 3.7%; Score 40.8; DB 13; Length 983;
 Best Local Similarity 53.0%; Pred. No. 2.7;
 Matches 87; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
 QY 287 tgcgcgtgcgcgtcttcacgtatcggtaccgacgacgtatcggtgcgttcggtg 346
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 732 GCCGAGACGGGAGCTGGCCCCCGTCCGCGCCGCGCTGTGGGGTTATTACCGT 673
 QY 347 cactcttataatcaaggcaggtgcgtggcgctgcaggtcttcgtgggtgtttctta 406
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 672 TGTATTATTATCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 613
 QY 407 ttgatctccagataggtcagttcttgggtgtggtgagcagttgtt 450
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 612 TTGTCGCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 569
RESULT 10
CNS03H6V/C
LOCUS CNS03H6V 970 bp DNA GSS 17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence r7 end of clone 026011 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL243904
VERSION AL243904.1 GI:7964916
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 970)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 970)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL REFERENCE Unpublished
AUTHORS 3 (bases 1 to 970)
 Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.
FEATURES location/Qualifiers
 source
 1..970
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG026AH06LP1-end : T7"
BASE COUNT 311 a 298 c 194 g 152 t 15 others
ORIGIN
 Query Match 3.6%; Score 40.2; DB 13; Length 970;
 Best Local Similarity 45.9%; Pred. No. 4;
 Matches 130; Conservative 2; Mismatches 151; Indels 0; Gaps 0;
 QY 219 tgcaggtgcttcaggtgtctccgctgtcttcttggtattacggagcctgcgattcttcgtgt 278
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 709 TCGTCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650
 QY 279 gaaccttcgcctgcgcgtgccttccttcacgtatcggtacccgagctatcggtggcgc 338
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 649 TGTTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 590
 QY 339 ttgattgacacttttaataatcaaggcaggttcggtggcgctgcaggtttcttcgtgtgt 398
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 589 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 530
 QY 399 tgtttctattgatgctccagataggtcatgttcttcgtgtgtgtgcagttgttacccttctt 458
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 529 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
 QY 459 catcgcatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 501
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 469 TGTTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 427
RESULT 11
BE997957
LOCUS BE997957 606 bp mRNA EST 06-OCT-2000
DEFINITION EST429680 GVSN Medicago truncatula cDNA clone pgVSN-889, mRNA sequence.
ACCESSION BE997957
VERSION BE997957.1 GI:10698233
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 606)
AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P., Town, C.D., Bowman, C.L., Craven, M.B., Cho, J. and Fraser, C.M.
TITLE ESTs from senescent nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: vance004@maroon.tc.umn.edu
 University of Minnesota name: M271935e TIGR sequence name: MTKAK05TK More information is available at:

THIS PAGE BLANK (USPTO)